DOCTOR OF PHILOSOPHY

New approaches to studying the GPI biosynthesis pathway in T. brucei uncovering the missing links

Ji, Zhe

Award date:
2021

New approaches to studying the GPI biosynthesis pathway in *T. brucei*: uncovering the missing links

ZHE JI

Division of Biological Chemistry and Drug Discovery

University of Dundee

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

2021

CONTENTS

List of figures

List of tables

List of abbreviations

AAT	African Animal trypanosomiasis
aa	amino acid
ATP	adenosine triphosphate
BSD	blastacidin S deamidase
BSF 2T1 WT	bloodstream 2T1 wild type
BSF	bloodstream form
BSF WT	bloodstream form wild type
CATT	Card Agglutination Test for Trypanosomiasis
Cas9 WT	Cas9 wild type
CDC	Centers for Disease Control and Prevention
cKO	conditional knock-out
Con A-like domain	concanavalin A-like lectins domain
CPSF3	Cleavage and Polyadenylation Sepcificity Factor 3
CTD	C-terminal domain
DAPI	4’,6-diamidino-2-phenylindole
DM	decyl-β-D-maltopyranoside
DIG	Digoxigenin
DFP	diisopropylfluorophosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DPM1	dolichol-phospho-mannose synthetase
DTTSSP	3,3'-dithiobissulfosuccinimidyl propionate
E. coli	*Escherichia coli*
ER	endoplasmic reticulum
ERAD	endoplasmic-reticulum-associated protein degradation
EtNP	ethanolamine phosphate
EtNP-T2	Ethanolamine phosphate transferase-2
EtNP-T3	Ethanolamine phosphate transferase-3
ES	expression site
XICs	eXtracted Ion Currents
gDNA	genomic DNA
GPI GnT	GlcNAc transferase complex
GlcN-PI	glucosamine-PI

GPI	Glycosylphosphatidylinositol
TbdeAc2	GPI inositol deacylase 2
GPI MTI	GPI-mannosyltransferase I
GC	guanine-cytosine
HA	hemagglutinin
HRP	horseradish peroxidase
HAT	Human African trypanosomiasis
IP	immunoprecipitation
ISG65	invariant surface glycoprotein 65
JBαM	jack bean α mannosidase
Galβ1-3GlcNAc	lacto-N-biose
LECA	Last Eukaryotic Common Ancestor
LB	Lysogeny Broth
MT I	mannosyltransferase I
Tm	melting temperature
MISP	metacyclic trypomastigote invariant surface protein
MW	molecular weight
Galβ1-4GlcNAc	N-acetyl-lactosamine
NECT	nifurtimox-eflornithine combination therapy
NOG	n-octyl-beta-glucoside
NTD	N-terminal domain
TLCK	Nα-Tosyl-L-lysine chloromethyl ketone hydrochloride
PNGaseF	Peptide -N-Glycosidase F
PBS	phosphate buffer saline
PI	phosphatidylinositol
PI-PLC PI-PLD	phosphatidylinositol-specific phospholipase C and D
PLA2	Phospholipase A2
PCR	Polymerase Chain Reaction
pol-I	polymerase-I
PARP	procyclic acidic repetitive protein
PCF	procyclic form
PCF WT	procyclic wild type
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
RDTs	Rapid Diagnostic Tests

rRNA	ribosomal RNA
RNAi	RNA interference
RIT-seq	RNA interference target sequencing
RHD3	root hair defective 3 GTP-binding protein
SSC	saline-sodium citrate buffer
SP	signal peptide
sgRNA	single guide RNA
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulphate polyacrylamide
sVSG117	soluble-form variant surface glycoprotein 117
SILAC	stable isotope labelling in cell culture
TbDPM2	dolichol-phosphate-mannose synthetase 2
TetR	tetracycline repressor
TLC	Thin layer chromatography
TfR	transferrin receptor
TX-100	Triton X-100
T. brucei	*Trypanosoma brucei*
T. cruzi	*Trypanosoma cruzi*
TDB	trypanosome dilution buffer
UbCE	ubiquitin-conjugating enzyme E2
UTR	untranslated region
VSGs	variant surface glycoproteins
VSG	variant surface protein
WHO	World Health Organisation

Acknowledgement

I consider myself a very lucky person to have studied in the School of Life Sciences, University of Dundee alongside such fine folk. My PhD has been one of the best experiences in my life.

I would like to thank the following people, without whom I would not have made it through my doctors' degree. I wish to express my deepest gratitude to my supervisor, Prof. Sir. Michael Ferguson, who has given me the opportunity and the best guidance and support possible. I would like to thank Mike for his constant help and patience through my PhD, not only in the academia aspect but also in my personal life. His rigorous research attitude and critical scientific thinking will make me benefit my entire research career. I must also acknowledge my thesis committee members Mark Field, David Murray and Nicola Stanley-Wall, who were instrumental in guiding me towards conducting a PhD. I thank Alvaro Acosta for the input in the TbGPI3 project. I thank CSC and University of Dundee for funding my PhD.

My special thanks to Dr. Lucia Güther, without whose day-to-day lab supervision, the goal of the project would not be realised. Thanks for the input and discussions in almost all aspects of this thesis! So many thanks to Sam for introducing me to the world of molecular biology, from understanding basic principles to plasmid construction, his patient explanations to all my questions have made cloning such an interesting thing instead of pain in the neck. I would also like to thank Michele for his help of the data analysis for all my proteomics data, his in-time support is the key of the smooth progress of the project. I thank Rupa for giving me advice on microbiology and mass spectrometry, Jose for the general help in the lab. I would also like to thank the past members in the group, Fernando, Di Wu, Elda and Giulia for the general discussions.

I would also like to thank DH, MOA and MCF lab members, especially Joana and Norma for the training of using microscope and IFA image analysis, Eva for the helping with the RNAi and CRISPR Cas9 design and protocols, Ric for helping with the *in situ* tagging strategy, Richard for RT-qPCR analysis, Pax and Sandra for help and discussion on micro and molecular biology.

Big thanks to Douglas and the Fingerprint proteomics facility team for the excellent support on proteomics.

I would like to thank all the great friends I met during my study. My special thanks to my Brazilian sister Luciana, I feel so fortunate to share these fantastic four years with her. I would like to thank my best friends Jiayue, Sandra, Rupa, Juliana, Carla, Marta and Weiwei for your precious friendship and company, I would never forget our fun girls' nights! Thanks also to my wonderful friends and colleagues; Xingui, Natalie, Andre, Anna, Douglas, Melanie, Cat.M, Victoriano, Lindsay, Rachel, Martin, Guilherme, Mattia, Kwok-Ho, Erika, Ross, Beatrice, Sam, Matthew.

My sincere gratitude to my parents for their unconditional support and encouragement throughout my PhD study, without whom I would not be who I am now. I am also lucky to have my sweet little brother Kailai to always keep me in an optimistic mood.

Finally, I would like to thank my boyfriend, Haonan Xu, for his support, love and always being there, without which I could never make it.

Declaration

I declare that I am the author of this thesis; all references cited have been consulted by myself; the work of which this thesis is a record, unless specifically stated, has been done by myself and this work has not been previously accepted for a higher degree.

Zhe Ji

I confirm that Zhe Ji has performed the research described in this thesis under my supervision and has fulfilled the conditions of the relevant Ordinance and Regulations of the University of Dundee.

Michael A.J. Ferguson

Summary

Trypanosoma brucei is a protozoan that causes Human African Trypanosomiasis and nagana in cattle. The bloodstream form (BSF) of *T. brucei* has a dense surface-coat of variant surface glycoprotein (VSG) that acts as the first line of defence against the host immune system while the main cell surface glycoproteins of the procyclic form (PCF) are the procyclins. VSGs and procyclins are attached to the cell membrane through glycosylphosphatidylinositol (GPI) anchors. Despite its importance, the GPI biosynthetic pathway in *T. brucei* is not fully elucidated.

In this thesis, we have applied epitope tagging, chemical crosslinking and quantitative proteomics methods to look for new components of protein complexes of the GPI pathway. Three components of the pathway were tagged and immunoprecipitated with or without chemical crosslinking and the results were as follows:

For TbGPI12, the GlcNAc-PI de-N-acetylase, we discovered that, contrary to expectations, it is not complexed with TbMTI but with a different potential partner protein. We also optimised methodologies in this part of the study.

For the TbGPI3 component of GPI GnT, we identified its expected partner proteins TbGPI15, TbGPI19, TbGPI2, TbGPI1 and TbERI1 and also identified two new components of the complex, namely an Arv1-like protein (TbArv1) and a putative E2-ligase.

For the GPI inositol deacylase (TbdeAc2) that catalyses inositol deacylation of GPI intermediates, we showed that it is present in a ~240 kDa complex and, using both crosslinking and stable isotope in cell culture (SILAC) methodology, identified partner protein(s) encoded by genes *Tb927.11.11740* and/or *Tb927.11.11750*. Individual reverse pull-down experiments showed that only Tb927.11.11750 was directly associated with TbdeAc2. Our attempts to generate a null and conditional null mutant for *Tb927.11.11750* were not successful. However, we considered the possibility that Tb927.11.11750 (which contains a lipase motif) encodes a GPI PLA2 enzyme; an activity adjacent to TbdeAc2 in the pathway. We therefore used immunoprecipitated Tb927.11.11750 in an *in vitro* assay for GPI PLA2 activity using a radiolabelled and purified glycolipid A' substrate. However, no GPI-PLA2 activity could be detected under the conditions used. Taken together, we postulate that *Tb927.11.11750* encodes an essential product for the survival of BSF *T. brucei* and hypothesise that the Tb927.11.11750 protein may be required for stabilisation and/or regulation of TbdeAc2.

1 Introduction

1.1 *Trypanosoma brucei*

1.1.1 Phylogenetics: The position of kinetoplastids in the eukaryotic kingdom and the position of *T. brucei* in the kinetoplastids

Trypanosoma brucei gambiense and *Trypanosoma brucei rhodesiense* are protozoan pathogens of the genus Trypanosoma and order Trypanosamatida that cause Human African trypanosomiasis (HAT) or sleeping sickness in humans. They are closely related to *T. brucei brucei* that, together with *T. congolense* species and *T. vivax* cause Nagana in cattle in sub-Saharan Africa. With the exception of *T. vivax*, all of these African trypanosomes are transmitted to their mammalian hosts via a blood meal of tsetse fly (Glossina genus). In this thesis I have performed experiments with clones of *T. brucei brucei* (Lister strain 427), referred to as *T. brucei* from herein.

All of the aforementioned trypanosomatid organisms belong to the class kinetoplastidea. The kinetoplastids, named after the morphological feature of a kinetoplast (containing DNA) in their mitochondrion and close to the basal body of their single flagellum, include other human pathogen trypanosomatids, notably *T. cruzi*, the causative agent of Chagas' disease endemic to South and Central America, and the leishmania which cause a variety of sandfly-transmitted pathologies throughout the tropics and sub-tropics. The kinetoplastids belonging to the phylum Euglenozoa in the major taxon group Excavata (Fig. 1.1).

Understanding biological origin and historical relationships with other organisms can help our understanding of parasite adaptation in invertebrate vectors and the vertebrate hosts. Phylogenetic analyses can be used to compare biological variants, providing information for disease-causing organisms in a molecular epidemiological perspective (Hillis, 1997). With the advances of molecular biology in the last century, the molecular phylogenetic analyses for the kinetoplastids has become available (Stevens and Gibson, 1999).

In the current consensus tree, a complex Last Eukaryotic Common Ancestor (LECA) is predicted to precede many unicellular organisms (Padilla-Mejia *et al.*, 2021). Before this specific period, but after the divergence from the archaeal lineage, it is thought that a prokaryotic lineage acquired typical features of eukaryotic cells, including organelle like compartments, including flagellum and endomembrane system, a cytoskeleton and

the mitochondrion. *T. brucei* is a unicellular organism that branches from this unique position in this early stage of the evolutionary tree, where the LECA emerged. Thus, the molecular evidence strongly suggests that the eukaryotic *Trypanosoma* genus emerged early from prokaryotes.

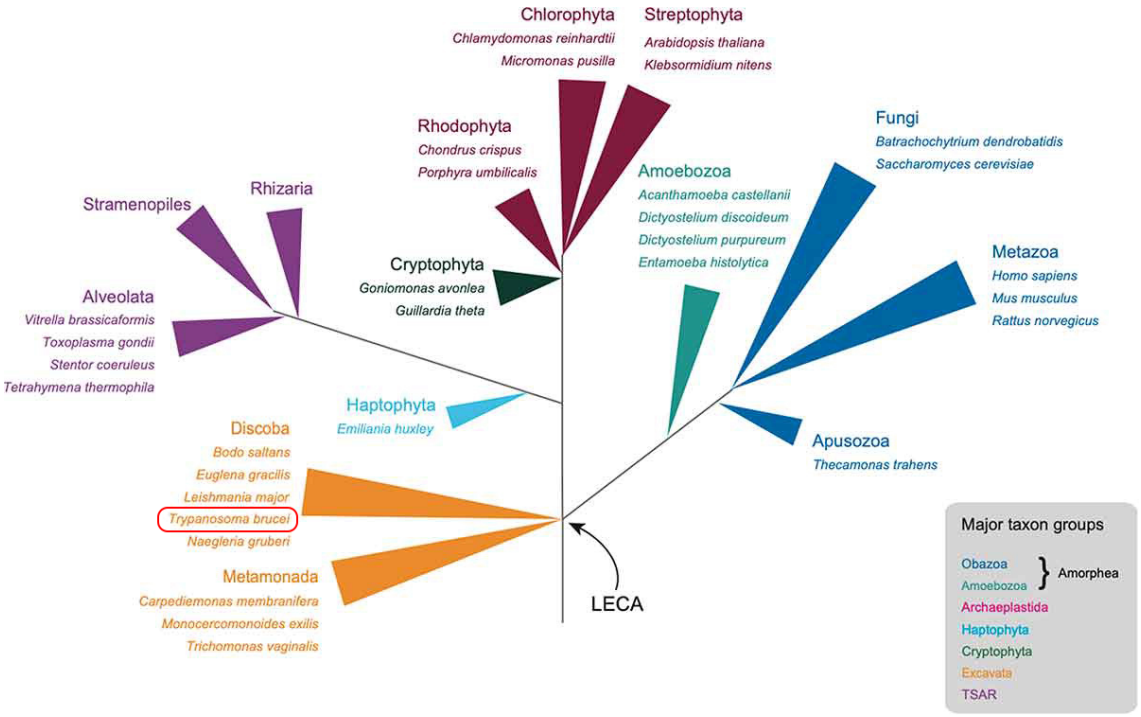

Figure 1.1 Schematic phylogenetic tree depicting the evolutionary relationships between the major eukaryotic taxa.

These taxon groups are annotated and colour-coded, while the *Trypanosoma brucei* clade is highlighted with a red rectangle in the tree. The tree is generated based on a recent study described in (Burki *et al.*, 2020). Image is taken from (Padilla-Mejia *et al.*, 2021).

Small-subunit ribosomal RNA (rRNA) sequences (Sogin, Elwood and Gunderson, 1986) or mitochondrial rRNA sequences (Lake *et al.*, 1988) were originally chosen to understand the relationships between kinetoplastid species. Using rRNA gene sequences, it was postulated that the evolutionary history of these eukaryotic parasites has connections with the origin of their insect hosts (Fernandes, Nelson and Beverley, 1993). It was also proposed that vertebrate parasitism arose multiple times in the trypanosomatids, with the evolution of trypanosomes is influenced more by their adaption to the invertebrate vectors than adaption to vertebrate hosts (Lukeš *et al.*, 1997). The most recent analysis, based on 18S rRNA sequences, places *T. brucei* in a distinct clade (Fig 1.2) (Borges *et al*., 2021).

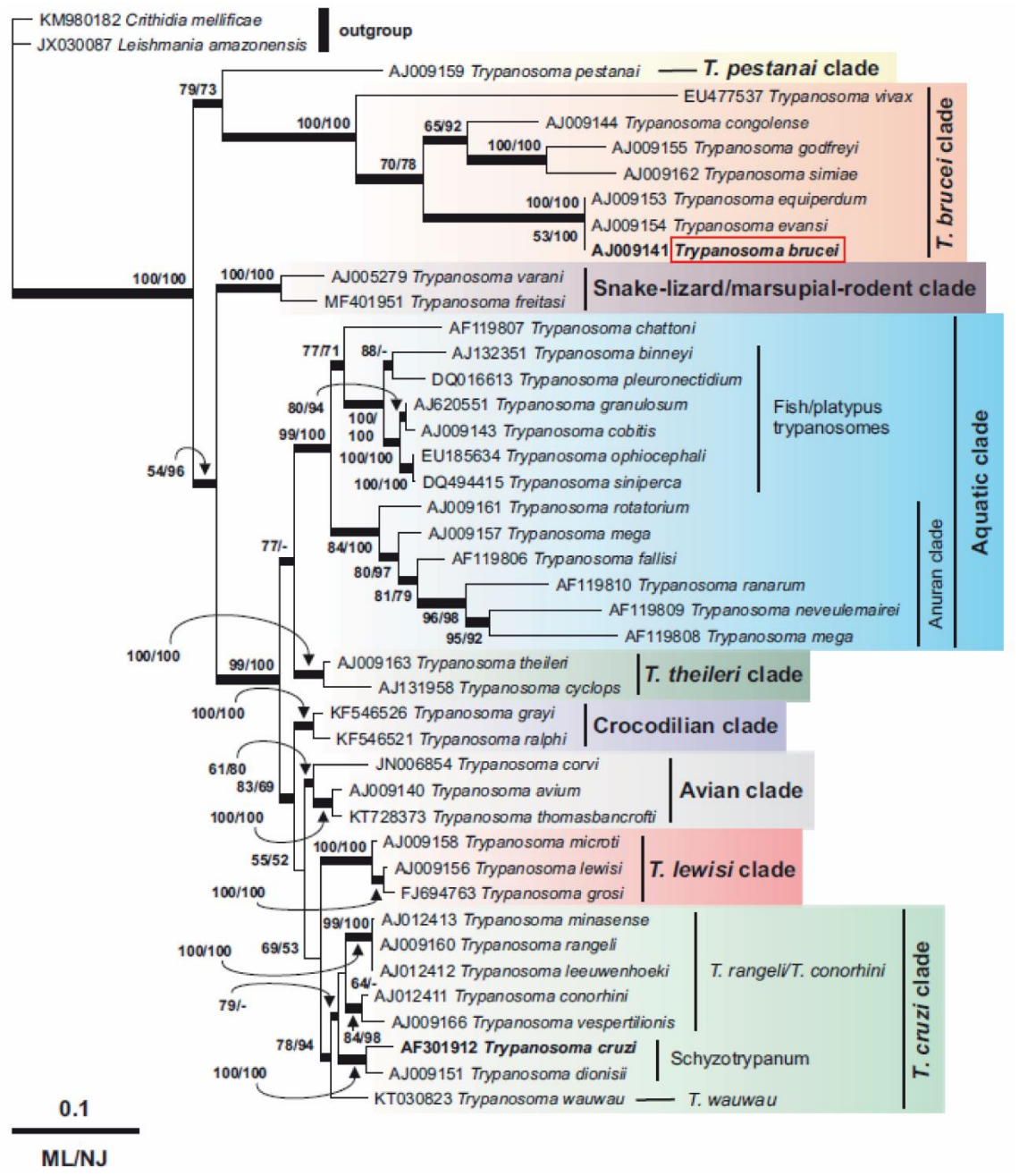

Figure 1.2 Phylogenetic trees based on bootstrapped maximum parsimony analysis of the 18S rRNA gene. *T. brucei* clade is highlighted in the red rectangle. Bootstrap values from 100 pseudo-replicates, mapped at the appropriate internodes, are from maximum likelihood-(ML) and neighbor-joining-analyses. Image is taken from (Borges *et al*., 2021)

1.1.2 Human African trypanosomiasis and Nagana

HAT or sleeping diseases in humans is caused by two subspecies of *Trypanosoma brucei*: *T. b. rhodesiense* causing acute African trypanosomiasis endemic in Western and Central Africa, and *T. b. gambiense* causing chronic African trypanosomiasis

endemic in Eastern and Southern Africa. Currently, this sub-species causes more than 95% of reported HAT cases. The two subspecies are morphologically indistinguishable. After infection, trypanosomes enter the haemo-lymphatic stage as they multiply in subcutaneous tissues, blood and lymph. The patients' symptoms in this stage include fever, headaches, enlarged lymph nodes, joint pains, and itching. The parasites can also cross the blood-brain barrier leading to the infection of the central nervous system. This stage is called the neurological stage. The symptoms for the patient in this stage are more evident with the most characteristic feature, which is the disturbance of the sleep cycle.

Trypanosomiasis caused devastating epidemics in the 20th century. With sustained control efforts over the past 20 years, the number of reported cases dropped below 10,000 for the first time in 50 years, and in 2020 led to a historically low number of 565 cases. This record has met the target of eliminating HAT as a public health problem, with fewer than 2000 cases per year (WHO, 2021) (Franco *et al.*, 2020).

The diagnosis for HAT alters depending on the patients' infection stage and varies between the diseases caused by different sub-species of *T. brucei*. In the case of *T. b. rhodesiense,* field diagnosis uses light microscopy to visualise the parasites on the infected individuals' thin or thick peripheral blood smear. It is typical for the patients to harbour high blood parasite levels with this infection (Kennedy, 2006). Primary screening for *T. b. gambiense* infection uses serological methods. The Card Agglutination Test for Trypanosomiasis (CATT) has been wildly adopted for disease screening due to its quick and straightforward testing features. However, the CATT methodology has several limitations, such as the high frequency of false-positive results and limited sensitivity (Kennedy, 2013). Therefore, this methodology is usually conducted in a region considered to have a high prevalence of disease (Kennedy and Rodgers, 2019).

Some advanced molecular techniques may improve the sensitivity and accuracy of the testing results. For example, Polymerase Chain Reaction (PCR) has been tested in HAT; however, it has been proved to have enormous technique-related issues and therefore not practical in the African region.

More recently, alternative effective diagnostic devices have been developed, called Rapid Diagnostic Tests (RDTs). One of the RDTs originated from our lab in collaboration with Professor Mark Carrington at Cambridge, an immunochromatographic test. To establish this test, antigens recognised specifically by HAT patients were first identified by proteomics and the four most promising were

expressed as recombinant proteins and tested by ELISA to compare immunoreactivities between *T .b. gambiense* infected and non-infected sera pools and individual sera. One of these, the extracellular domain of a *T. b. gambiense* invariant surface glycoprotein 65 (ISG65) gene was prototyped by itself (Sullivan *et al* 2013) and together with native soluble-form variant surface glycoprotein 117 (sVSG117) (Sullivan *et al.*, 2014) and compared with other native and recombinant antigens (Sternberg *et al.*, 2014). This strategy led to the development of the commercial BIOLINE HAT 2.0 test which uses recombinant ISG65 and a recombinant VSG that has high sensitivity and specificity. It was recently announced that 450,000 copies of this RDT will be donated by global healthcare company Abbott to scale up the testing for sleeping sickness from 2021.

For HAT treatment, World Health Organisation (WHO) has updated the pharmacological therapy for *gambiense* HAT in 2020 with substantial changes for clinical practice, including the oral fexinidazole monotherapy into the first-line treatment (Lindner *et al.*, 2020). Before this new treatment, previous pharmacological therapy for *gambiense* HAT depended on a few old drugs which have toxicity liabilities. Here is a summary of current treatment guidelines from WHO (Lindner *et al.*, 2020): Drug treatment for HAT is dependent on the stage of the disease for both HAT variants. For early-stage *T. b. gambiense* infection, fexinidazole or pentamidine are the first-line drug treatments with the recommendation of using fexinidazole in most cases. However, pentamidine is the only approved treatment for patients under age six or whose weight is below 20 kg. For the second-stage *T.b. gambiense* infection, fexinidazole or nifurtimox-eflornithine combination therapy (NECT) are used.

As for the treatment for the early-stage of *T. b. rhodesiense*, suramin is given by intravenous administration. This treatment is proven to be effective but has several potential side effects. Drug treatment for the second-stage disease is more problematic with the only treatment currently available being intravenous melarsoprol. This drug, although very effective, is very toxic, with a fatality rate of around 5-9% (Kennedy and Rodgers, 2019).

Fexinidazole has proven to be effective for the treatment of *T. b. gambiense* with a 10-day oral course (Lindner *et al.*, 2020). However, there is a high risk of relapse after the treatment, and this relapse may occur late after the treatment, up to 12-24 months. Therefore, further studies need to be conducted to solve these risk factors, and more clinical trials need to be carried out for its use in children under six and patients whose weight is less than 20 kg. In conclusion, fexinidazole has a prospect for simplifying the

treatment of *gambiense* HAT. It will be easier to utilise in poorly resourced areas where the standard pharmacovigilance systems are challenging to set up.

Recently, the mode of action of two new benzoxaboroles drugs that are now in clinical trial were revealed with targeting the Cleavage and Polyadenylation Sepcificity Factor 3 (CPSF3) (Wall *et al.*, 2018). These benzoxaboroles could be orally taken which will provides an additional patient-friendly therapies for both sleeping sickness and Nagana.

As mentioned, the goal of WHO to eliminate sleeping sickness as a public health problem by 2020 has been achieved. However, there are still several challenges that should be taken into consideration: i: Resurgence of HAT, as were recorded in the 1960s even after it had appeared to have been brought under control. ii: The frequency of asymptomatic patients is still unknown, and there is a reservoir of trypanosomes being identified in the skin and extravascular tissues in asymptomatic humans, which will all be an issue for the elimination of HAT as a public health issue. iii: Some cases occur in the rural area of sub-Saharan Africa, which poses a problem for identifying and recording infections. iv: The current drug treatment for *rhodesiense* HAT is not adequate for the second-stage disease. Melarsoprol treatments are extremely painful and bring many side effects with even consequences of post-treatment fatality (Kennedy and Rodgers, 2019).

African Animal trypanosomiasis (AAT) is caused by subspecies of the *Trypanosoma* genus, mainly *T. congolense*, *T. vivax* and *T. brucei brucei* subspecies. In cattle, this disease is called Nagana. The animal trypanosomiases, especially Nagana, are a significant economic burden that affects many rural areas of Africa. In addition, animals (mainly cattle) can act as a reservoir for the human infectious parasites, especially *T. b. rhodesiense*. The fact that cattle are reservoirs of *T. b. rhodesiense* posts more challenges to control this HAT variant. Although local control has been successful, it is probably not a realistic proposition to eliminate the tsetse fly vector, given that there is no evidence of diminishing regions that the tsetse fly infests over the last 100 years.

In conclusion, the surveillance, diagnosis and drug treatment for *gambiense* HAT and early-stage *rhodesiense* HAT are now quite well established. However, the drug treatment for second-stage *rhodesiense* HAT with the only first-line drug melarsoprol that causes severe side effects are far from adequate. Therefore, developing more reliable and affordable diagnostic staging biomarkers that could discriminate the early-stage disease from late-stage disease is an alternative way to helping providing sufficient treatment in time and avoid the use of toxic drugs for late-stage disease. Meanwhile, more effective, safer, and easier administration of new drugs need to be

developed to achieve better treatment for *rhodesiense* HAT. As for animal trypanosomiasis, more obstacles remain to be tackled, such as new treatment methods and control of the transmission. Recently, a new study has broadened the field's view on how to control the disease, introducing *in vivo* protective immunity against invariant *Trypanosoma vivax* vaccine antigen (Autheman *et al.*, 2021). The authors managed to find a single recombinant flagellum membrane protein for vaccination that could induce long-lasting protection. The monoclonal antibodies to this protein could clear the parasitaemia in the mouse model, including the significant role for a complement. Their study shows that protective vaccines against trypanosome infections can be achieved and provides a vaccine candidate for controlling and curing *T. vivax* trypanosomiasis.

1.1.3 Life cycle of *Trypanosoma brucei*

T. brucei undergoes a complex life cycle between its teste fly vector and mammalian hosts. It is a unicellular protozoan with six main cellular morphologies dependent on different cellular organisations, like the position of the kinetoplast relative to the posterior end of the cell. The life cycle of *T. brucei* is summarised in (Fig 1.3). The metacyclic trypomastigotes are co-injected with saliva into the mammalian host's skin tissue, entering the lymphatic system and bloodstream. Here, they differentiate into long slender bloodstream trypomastigotes, and the proliferative life cycle stage starts with replication of the parasites by binary fission. When the parasitaemia in the mammalian host increases, some cells develop into a stumpy form and are arrested in the G1/G0 phase, ready to be taken by the insect vector (McKean, 2003). The short stumpy trypomastigotes that tsetse fly takes then enter the vector's midgut and differentiate to procyclic trypomastigotes in which stage the binary fission takes place for proliferation. During the migration to the salivary glands, the parasite transforms into epimastigotes and continue multiplication by binary fission. Inside the salivary glands, some epimastigotes will differentiate into metacyclic trypomastigotes. Studies (Peacock *et al.*, 2014) have also shown the identification of haploid gametes in the insect stage. Interactions and cell fusions are detected between these gametes, indicating the sexual reproduction process in *T. brucei*. The complete process in the tsetse fly takes around 20 days.

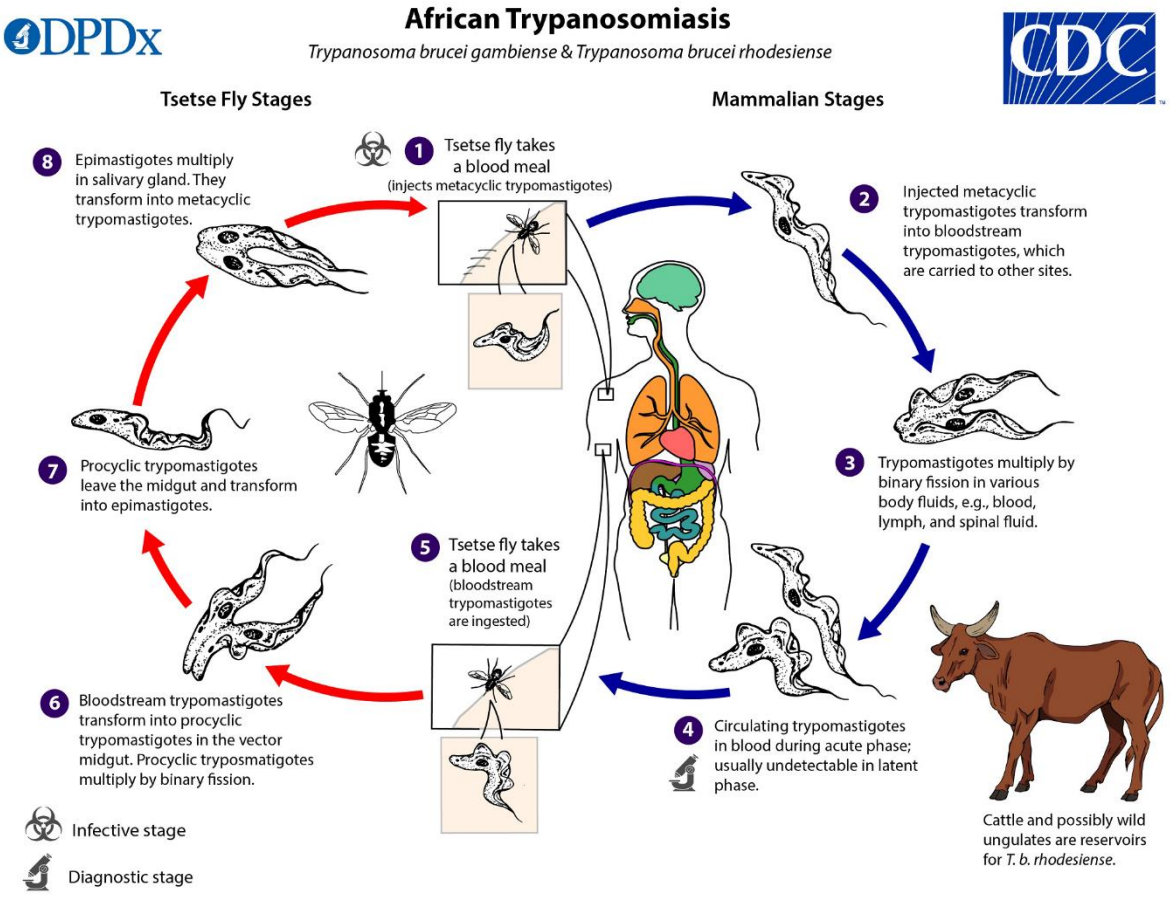

Figure 1.3 Life cycle of *Trypanosoma brucei*.

The transmission of *T. brucei* starts with a bite on the mammalian host by a tsetse fly. After the metacyclic trypomastigotes enter the mammalian host, these cells differentiate to dividing slender bloodstream-form and then some of these to non-dividing stumpy form. At this stage, they can be taken up by the insect vector, where the differentiation to procyclic form takes place in the fly midgut. Asymmetric division gives rise to the short epimastigote form that migrates and attaches to the salivary gland. Here the cells proliferate and mature to the infectious metacyclic form. The image is taken from CDC (Centers for Disease Control and Prevention).

1.2 Surface structures of *Trypanosoma brucei*

1.2.1 Overview of the known and predicted GPI-anchored glycoproteins of *T. brucei*

The glycosylphosphatidylinositol (GPI) anchor is an almost universal eukaryotic protein attachment mechanism. In *T. brucei*, there are a lot of proteins with different functions that are GPI anchored. With the help of the GPI anchor prediction tool (Eisenhaber, Bork and Eisenhaber, 1999) applied to the predicted proteins of *T. brucei*, a list of 155 proteins (zenodo:10.5281/zenodo.5338935) predicted to be GPI-anchored was

generated. For some of these there are experimental data supporting the prediction. The GPI prediction algorithm looks for sequences that contain an N-terminal signal peptide (to enter the lumen of the ER) and for a GPI attachment signal peptide with the following features: The amino acid to which GPI is attached is termed the ω site amino acid (Gerber, Kodukula and Udenfriend, 1992), which is usually a relatively small and polar residue (Ala, Asn, Asp, Cys, Gly, or Ser). The GPI attachment signal peptide contains 20-30 amino acid and starts from the ω+1 amino acid. The repertoire of amino acids in the ω+1 and ω+2 sites are similar to that for the ω site and the algorithms sometimes struggle to correctly pick the ω site for this reason. These are followed by 5-10 relatively polar amino acids and then a run of 15-20 hydrophobic amino acids to the predicted C-terminus. This C-terminal signal peptide is recognised by the GPI transamidase (which will be introduced later), and cleavage occurs between the ω and ω+1 amino acids.

1.2.2 Main surface structure of BSF *T. brucei*-GPI anchored VSGs and antigenic variation

The bloodstream form (BSF) of *T. brucei* produces a dense coat of GPI anchored (Ferguson et al Science 1988) variant surface protein (VSG) which consists of ~10 million molecules that accounts for at least 95% of cell-surface proteins (Ziegelbauer and Overath, 1992) (Grünfelder *et al.*, 2002). This dense layer acts as the first defence and protects the parasite from the innate immune system of the host. The expression of a VSG coat on *T. brucei* starts on the metacyclic trypomastigotes (from a specific metacyclic VSG gene repertoire) when it resides in the salivary glands of the teste fly (Rotureau and Van Den Abbeele, 2013). Upon differentiation to the BSF a wider repertoire of VSG genes is available for expression (Cross *et al*., 2014).
Although efforts have been put into understanding the molecular mechanisms of how VSGs defend the against the host's immune system for decades, it is only recently years that the molecular organisation of the intact VSG coat has been proposed (Bartossek *et al.*, 2017). The first crystal structures of two VSGs, MITat1.2 and ILTat1.24, were solved three decades ago with a resolution of 6 Å (Freymann *et al.*, 1984) (Metcalf *et al.*, 1987). These early structures have provided essential basic understandings of the structures of VSGs as, generally, VSG homodimers of 100-120 kDa with a conserved coiled coil tertiary architecture, despite low similarity in primary sequence. However, these and later VSG crystal structures (Pinger et al., 2018) only provide structures for the N-terminal domain (NTD) of VSGs. The crystallisation of intact VSGs purified by

standard purification methods leads to crystals containing only the NTDs. The lack of C-terminal domain (CTD) of VSG through purification provides evidence that the unstructured region, Linker 1(L1) (Schwede *et al.*, 2015), connecting the NTD and the CTD is likely accessible and sensitive to proteases. With continuous efforts on solving the structures of VSGs, the NMR solution structures of CTD of VSGs MITat1.2 and ILTat1.24 were determined in 2005 (Chattopadhyay *et al.*, 2005) and 2008 (Jones *et al.*, 2008), respectively. In 2017, another NMR solution structure of the CTD of VSG M1.1.a was added to the database. The comparison of these three different CTDs of different VSGs highlights that they all share conserved core structure despite the differences in the primary sequences. Models for the complete structure of VSG M1.1 were determined by fitting the x-ray crystallographic NTD and solution NMR CTD structures into small angle X-ray scattering (SAXS) envelopes of intact VSG. These suggested that it could adopt two main groups of conformations with evident flexibility of the Linker 1 of VSGs (Fig 1.4). The authors of this paper proposed that the conformation of VSGs is greatly affected by the environment when they are bound to the cell membrane. The NTDs and CTDs of VSGs structure can also be grouped into two types for NTDs and six types for CTDs (Schwede *et al.*, 2015).

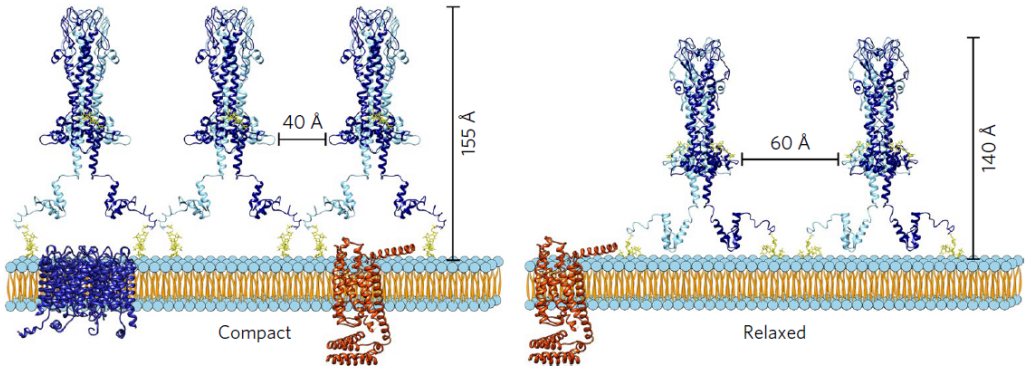

Figure 1.4 Proposed two packing methods of GPI-anchored VSGs on the cell surface. Illustration of two forms of VSGs surrounded in two different environments with tightly packed VSGs at left and the relaxed version at right. Image is taken from (Bartossek *et al.*, 2017).

T. brucei has a reservoir of up to 2000 VSG genes and gene-fragments, which make up of 30% of its genome (Horn, 2014). Of these, about 20% are competent to encode full-length functional VSGs (Cross *et al*., 2014) (Only one out of these VSG genes is activated and transcribed at a time by polymerase-I (pol-I) (Horn, 2014) as part of the

polycistronic transcription unit at the expression site (ES) (Vol *et al.*, 1982). In a natural infection, only a small number of "early" VSG genes are being expressed by the parasite population and the host develops a specific immune response to these VSGs, and the parasitaemia is attenuated by the antibody response. However, some parasites switch to expressing another VSG gene in a different telomeric expression site, Further, telomeric expression sites can change VSG genes by homologous recombination due to the 70 bp repeats and the conserved 3'-UTR flanked with these VSG genes (Fig 1.5). Antigenic variation of VSGs allows this parasite to evade the adaptive immune system.

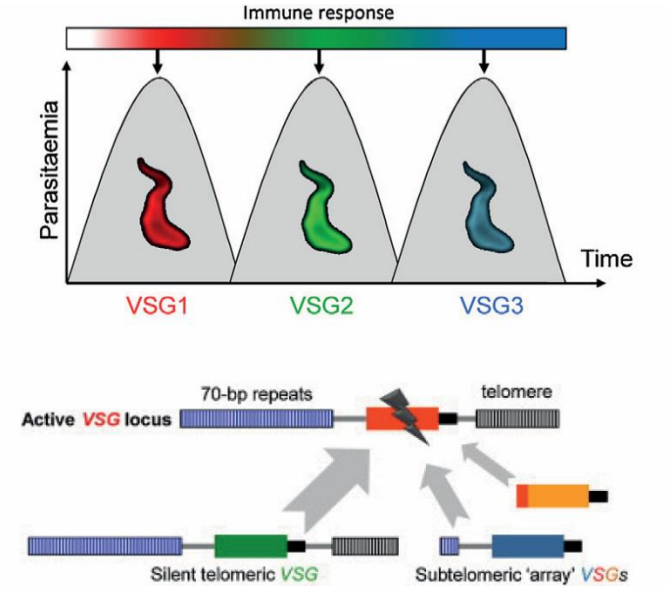

Figure 1.5 VSG expression and antigenetic variation.

Simplified scheme of successive immune responses due to the antigenic variation of VSGs. The image is taken from (Horn, 2014). DNA recombination is central to VSG switching.

In addition to this fundamental mechanism of antigenic variation by VSG gene switching, a recent structural study of VSG3 also shows that post-translational modification on the surface of VSGs can add to the variability of VSGs (Pinger *et al.*, 2018). Heterogeneous O-glycosylation with zero to three hexose was present on VSG3 and other VSGs. This O-glycosylation increased the parasites' virulence by interfering with the host's generation of immunity (Fig 1.6).

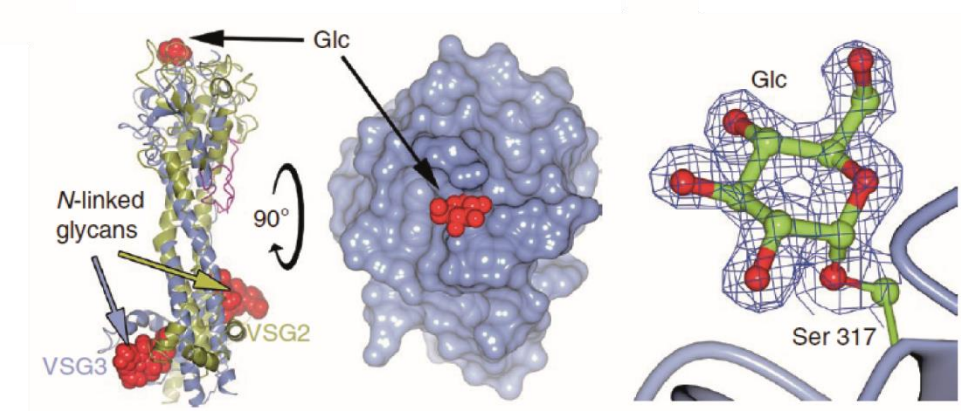

Figure 1.6 Identification of O-linked glycosylation of VSG3 through crystallisation.
Left: Structure of VSG3 (blue) and VSG 2 (gold) with identifying a red space-filling α -glucose moiety attached to Ser 317 of VSG3. Right: A 1.4-Å-resolution electron density map showing the linkage between the modelled glucose ad the Ser 317 representing the main protein chain. Image is taken from (Pinger *et al.*, 2018)

1.2.3 Other known and predicted BSF GPI-anchored glycoproteins

In *T. brucei*, there are also other proteins that are known to be attached to the BSF cell membrane through GPI anchors. Several proteomics analyses that focused on the cell surface proteins of *T. brucei* (Shimogawa *et al.*, 2015) (Zoltner *et al.*, 2015)have also expanded the datasets for analysing potential GPI anchored proteins that are present on the parasites' plasma membrane. For example, an important surface molecule that has been shown structurally to possess a GPI anchor is the ESAG6-subunit of the BSF transferrin receptor (TfR) (Mehlert and Ferguson, 2007), although surprisingly the GPI anchor itself can be replaced by a transmembrane domain to still yield a functional TfR in BSF *T.brucei* (Kabiri and Steverding, 2021).

The BSF haptoglobin-haemoglobin receptor is also predicted to be a GPI-anchored protein in *T. brucei* with an N-terminal signal peptide and C-terminal GPI addition signal peptide (Vanhollebeke *et al.*, 2008) (Lane-Serff *et al.*, 2014). This protein plays a role in mediating high-affinity binding with haptoglobin-haemoglobin for the uptake and incorporation of heme into intracellular hemoproteins. It also acts as a receptor of host serum trypanolytic factor-1 that protects humans from *T. brucei brucei* infections. The factor H receptor is another protein with the predicted N-terminal signal peptide and C-terminal GPI addition sequences. This factor H receptor can bind mammalian factor H with the inhibitory domain free to inactivate complement C3b deposited on the trypanosome surface (Macleod *et al.*, 2020). It has been shown to be a virulence factor

for the transmission of trypanosomes, with the highest expression level in the developmental stages transmitted to the tsetse fly vector and those exposed to blood meals in the tsetse gut. Other examples like the epimastigote stage-specific surface protein BARP glycoprotein (Urwyler *et al.*, 2007) and the metacyclic trypomastigote invariant surface protein (MISP) (Casas-Sánchez *et al.*, 2018) have also been predicted to be GPI-anchored proteins.

1.2.4 Main surface structure of PCF *T. brucei* – procyclins

The major surface glycoproteins of the tsetse mid-gut dwelling procyclic form (PCF) of the parasite are the procyclins and these are also a GPI-anchored proteins. The change of trypanosome surface coat happens when the stumpy forms differentiate to the procyclic forms where the VSGs coat exchanged to the procyclin coat (Metcalf *et al.*, 1987) (Matthews and Gull, 1994), which has two main forms. The procyclins, which were also previously called procyclic acidic repetitive protein (PARP), comprises about 1% PCF trypanosome protein and was first isolated in 1989 (Clayton and Mowatt, 1989) by purification methods involving cell fractionation, ion exchange and concanavalin A-Sepharose affinity chromatography. The structures of two forms of PARP/procyclin were determined from material purified from long-term cultures of *T. brucei* PCF. These were purified by solvent extraction and hydrophobic interaction chromatography, which has become the preferred route to procyclin purification. This procedure yielded both Glu-Pro (EP) repeat domain-containing procyclin, which contains an N-glycosylation site, and a Glu-Pro-Pro-Thr (GPEET) repeat domain-containing procyclin that lacks an N-glycosylation site (Treumann *et al.*, 1997). The GPEET-procyclin was later also shown to be hyper-phosphorylated (Treumann *et al.*, 1997).

Later, the procyclins were analysed *in vivo* during the establishment of midgut infections in the tsetse flies (Acosta-Serrano *et al.*, 2001) by mass spectrometry. The procyclins are the products of a small multigene family, and the analysis showed the expression of different procyclins in an orderly manner. In the early infection, GPEET2 is the only procyclin detected. As the infectious days' increase, GPEET2 is gradually replaced by several isoforms of glycosylated EP forms. Also, the N-terminal domains of all procyclins are sensitive to proteolysis in the tsetse fly. In contrast, the C-terminal domain appears to be protease-resistant, indicating that the C-terminal domain

containing the polyanionic amino acid repeats may protect the parasite surface from digestive enzymes in the tsetse fly gut.

1.3 GPI anchor

1.3.1 GPI anchor structures – common core structure and species and tissue-specific lipid, carbohydrate side-chain and phosphodiester substituent variations

Together with proteins mentioned previously, dozens of GPI-anchored proteins have been identified and several hundred postulated in many eukaryotes, suggesting that protein-lined GPI-anchors ubiquitous among eukaryotes. Proteins containing a GPI anchor play roles in a wide variety of biological processes, including cell-cell recognition, cell signalling, cell surface enzymology and cell protection.
Thus far, GPI anchor structures have been completely or partially solved for four *T. brucei* VSGs (Ferguson *et al.*, 1988) (Mehlert *et al.*, 1998) (Mehlert *et al.*, 2010) (Treumann *et al.*, 2010) , the TfR (Mehlert *et al*., 2012) and the procyclins (Field *et al*., 1991) (Treumann *et al.*, 1997). As for the structure of GPIs, research on *T. brucei* was the first to yield methodologies to delineate the steps of GPI biosynthesis that were subsequently applied to mammalian cells and yeast (Masterson *et al.*, 1989) (Masterson *et al.*, 1990) (Menon *et al.*, 1990). GPI structures have also been completely or partially solved for several other eukaryotic species both unicellular (*T. congolense, T. cruzi*, *Leishmania major*, *Plasmodium*, *Toxoplasma*, *Neospera*, *Paramecium*, *Saccharomyces*, *Aspergillus*) and multicellular (*Dictyostelium*, *Pyrus, Torpedo*, as well as mammalian (rat, bovine, human, hamster, pig))(Nett *et al.*, 2009) (Ferguson *et al*., 2017)
The analysis of the protein GPI anchor structures shows that GPI anchors have a conserved core structure of ethanolamine-P-6Manα1-2Manα1-6Manα1-4GlcNα1-6*myo*-inositol-1-P-lipid. The ethanolamine phosphate (EtNP) acts as a bridge to the C-terminal amino acid of a mature protein, making a amide bond to the C-terminal amino acid α-carboxyl group. The lipid moiety is generally the sole means of attachment of such proteins to the membrane (Fig 1.7). However, outside of *T. brucei*, not all GPI anchors have a diacylglycerol lipid component, alkyacyl-glycerol and ceramide lipids are also common, particularly in mammalian and yeast GPI anchors, resectively]. The structure shown in (Fig 1.7) is the general structure for all protein-linked GPI anchors. For BSF *T. brucei*, the R_1 and R_2 fatty acid chains are exclusively myristic acid (C14:0) due to the trypanosome-specific phenomenon of fatty acid remodelling of the GPI

precursor prior to exchange for the GPI anchor addition signal peptide (Masterson *et al*., 1990) (Güther *et al*., 1994), see (Fig 1.8). For the mature VSG and TfR GPI anchors, R_3 is OH, whereas the BSF GPI biosynthetic precursors are in dynamic equilibrium between R_3 being OH and an ester-linked fatty acyl group (Güther and Ferguson, 1995), which is mostly palmitate (C16:0) (Güther *et al*.,1996) (Fig 1.8). The latter are referred to as being inositol-acylated. After the non inositol-acylated GPI anchor precursor (referred to as glycolipid A) is transferred to protein by the GPI transamidase complex in the ER, it is further processed in the ER and Golgi apparatus where the GPI anchor can be galactosylated at the R_6, R_8 and R_{10} positions (see Fig 1.8). The GPI anchor structure is different for PCF *T. brucei* (Treumann *et al*., 1997) such that: (i) R_1 at the *sn*-2 position of PI is OH with R_2 mostly stearic acid (C18:0); (ii) The inositol is acylated with predominantly palmitate acid at the R_3 position. (iii) The side chain modifications are extremely complex, containing at R_8 Gal and GlcNAc in the form of N-acetyl-lactosamine (Galβ1-4GlcNAc) and lacto-N-biose (Galβ1-3GlcNAc) repeats capped with α2-3 linked sialic acids (see Fig 1.8).

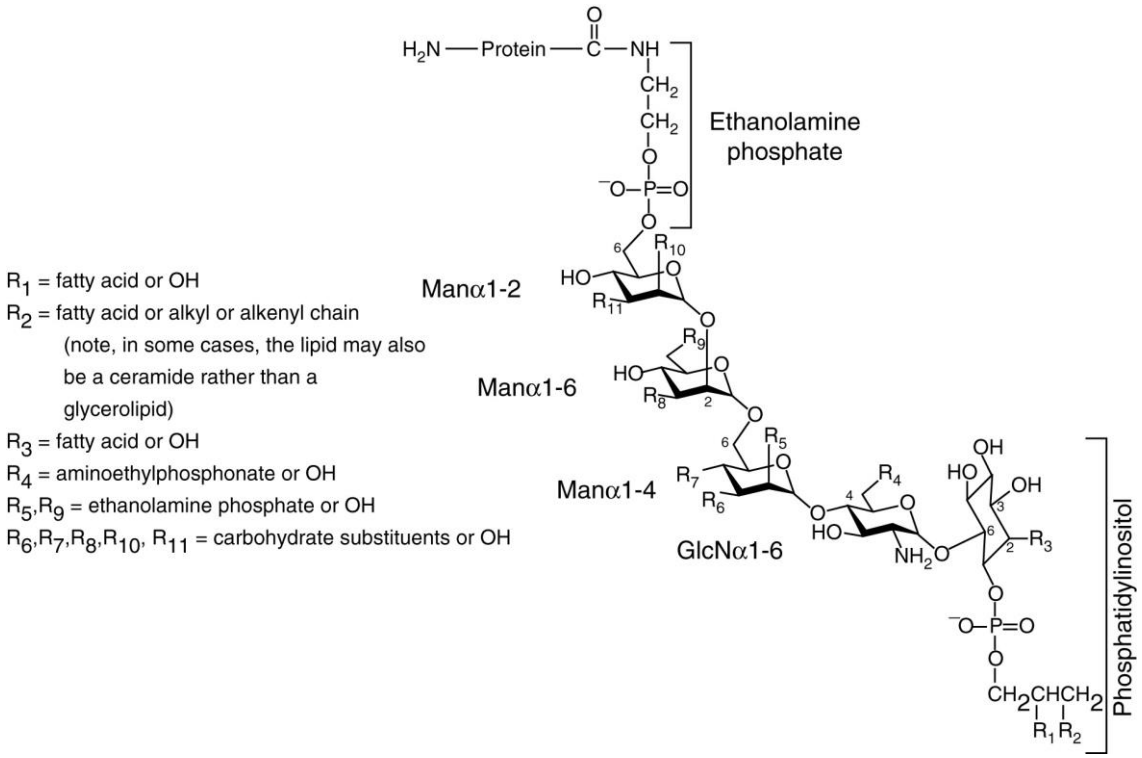

Figure 1.7 General structure of glycosylphosphatidylinositol (GPI) anchors attached to proteins.

They all share a common core consisting of ethanolamine-PO_4-6Manα1-2Manα1-6Manα1-4GlcNα1-6myo-inositol-1-PO_4-lipid. Various substitutions on different GPI anchors are indicated here as R. Image is adapted from (Ferguson *et al*., 2017).

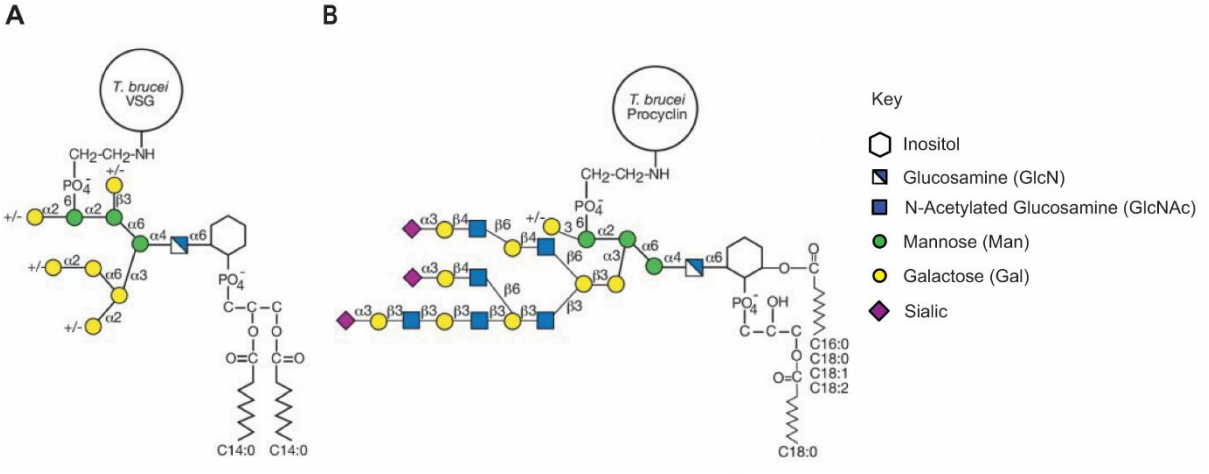

Figure 1.8 The structures of (A) the BSF *T. brucei* VSG GPI anchor (Ferguson *et al.*, 1988) and (B) a consensus structure of the PCF *T. brucei* procyclin GPI anchor (Treumann *et al.*, 1997) (Izquierdo *et al.*, 2009). Images taken from (Ferguson *et al*., 2017).

The chemical properties of GPI anchors are quite complex as they contain amide, glycosidic, phosphodiester and hydroxyester linkages between their various components. The determination of the GPI anchor structure was favoured by several chemical and enzyme cleavages of the GPI anchor. Several of these reactions were later applied to the detection of GPI anchors on GPI-anchored proteins and to the characterisation of GPI biosynthetic intermediates. The commonly used reagents, enzymes and corresponding reactions with products are illustrated in (Fig 1.9). Panel A shows two key chemical reactions of GPI anchor. Aqueous hydrogen fluoride treatment releases the GPI glycan from PI and the ethanolamine phosphate, allowing the released glycans to be permethylated and analysed by tandem mass spectrometry. Nitrous acid deamination of the glucosamine residue causes a highly selective cleavage of the glucosamine-inositol glycosidic bond, releasing the PI moiety. The liberated PI can be extracted by solvent partition and analysed by mass spectrometry. The general enzymatic digestions are shown in Fig 1.9 panel B. The digestion of a mature GPI anchor with jack bean α mannosidase (JBαM) does not liberate any free mannose from the structure as the terminal mannose is linked to the ethanolamine phosphate residue (Ferguson *et al.*, 1988). However, a GPI intermediate lacking the ethanolamine phosphate will be fully digested. Phospholipase A2 (PLA2) digestion will release the fatty acid chain at *sn*-2 position from the GPI anchor, which helps to distinguish different GPI anchor intermediates. Both phosphatidylinositol-specific phospholipase C and D (Fig 1.9 panel C) can release the lipid or phosphatidic acid portion from the GPI

anchor while the susceptible substrates for the reactions are different. PI-PLC cannot digest the GPI anchor if the inositol ring is acylated, while PI-PLD works on both inositol acylated and non-acylated GPI anchors.

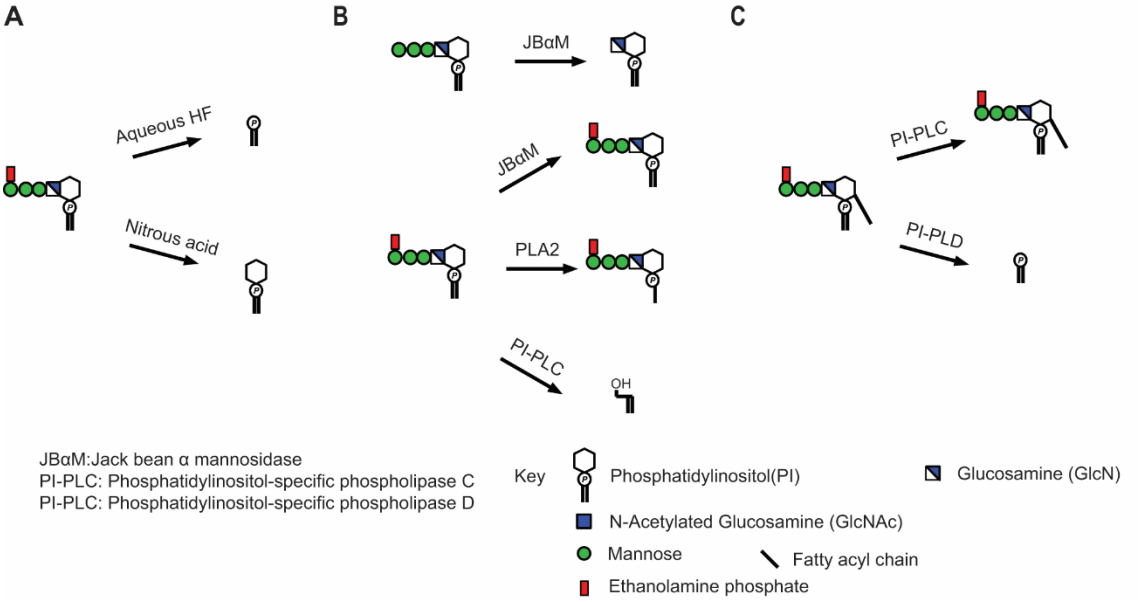

Figure 1.9 Digestion patterns of GPI anchors by chemical reagents or enzymatic reactions. A: Aqueous HF releases the PI portion from the GPI anchor while the nitrous acid deamination of the GlcN releases the PI component. B: JBαM cannot digest a GPI anchor intermediate where the terminal mannose is protected by ethanolamine phosphate. PLA2 can liberate the *sn*-2 fatty acid chain from GPI anchor. PI-PLC will release lipid portion from non inositol-acylated GPI anchors but not inositol-acylated GPI anchors D: PI-PLD will liberate phosphatidic acid from inositol acylated and non-acylated GPI anchors.

1.3.2 GPI anchor biosynthesis pathway in *T. brucei*

The biosynthesis of GPI anchor in *T. brucei* has been illustrated through a series elegant radiolabelling experiments using living cells and BSF and PCF trypanosome cell-free systems (Masterson *et al.*, 1989) (Masterson *et al.*, 1990) (Menon *et al.*, 1990) (Field *et al.*, 1991) (Güthee *et al.*, 1994) (Güther and Ferguson, 1995). The scheme of the GPI anchor biosynthesis pathway of BSF *T. brucei* is shown in (Fig 1.10). The pathway is located in the endoplasmic reticulum (ER) compartment of the parasite (Vidugiriene and Menon, 1994). The whole process starts in the cytosolic face of ER where N-acetylglucosamine is added to phosphatidylinositol (PI) following by the de-N-acetylation of GlcNAc-PI yielding to glucosamine-PI (GlcN-PI). The GlcN-PI is flipped

into the lumenal leaflet of ER where it is mannosylated sequentially with three mannoses using the donor dolichol-phosphate-mannose (DPM). After the addition of the first mannose, all GPI intermediates are in equilibrium between their inositol-acylated and non-acylated forms. The ethanolamine phosphate is attached to the third mannose from the donor phosphatidylethanolamine, after which the fatty acid re-modelling of the PI occurs involving the sequential exchange from longer chain fatty acids to myristic acids on the *sn*-2 and *sn*-1 positions. After the synthesis of the mature GPI anchor precursor (glycolipid A), proteins displaying a C-terminal GPI addition sequence have the glycolipid A precursor transferred to them by the GPI transamidase complex (Nagamune *et al.*, 2003) (Hong *et al.*, 2006). Following this, sidechain glycosylation in the ER and Golgi apparatus occurs and the processed GPI-proteins are transported to the plasma membrane via the flagellar pocket. Each step of the GPI anchor biosynthesis pathway is catalysed by an enzyme or a protein complex, the list of enzymes that are involved in this pathway is listed in (Table 1.1).

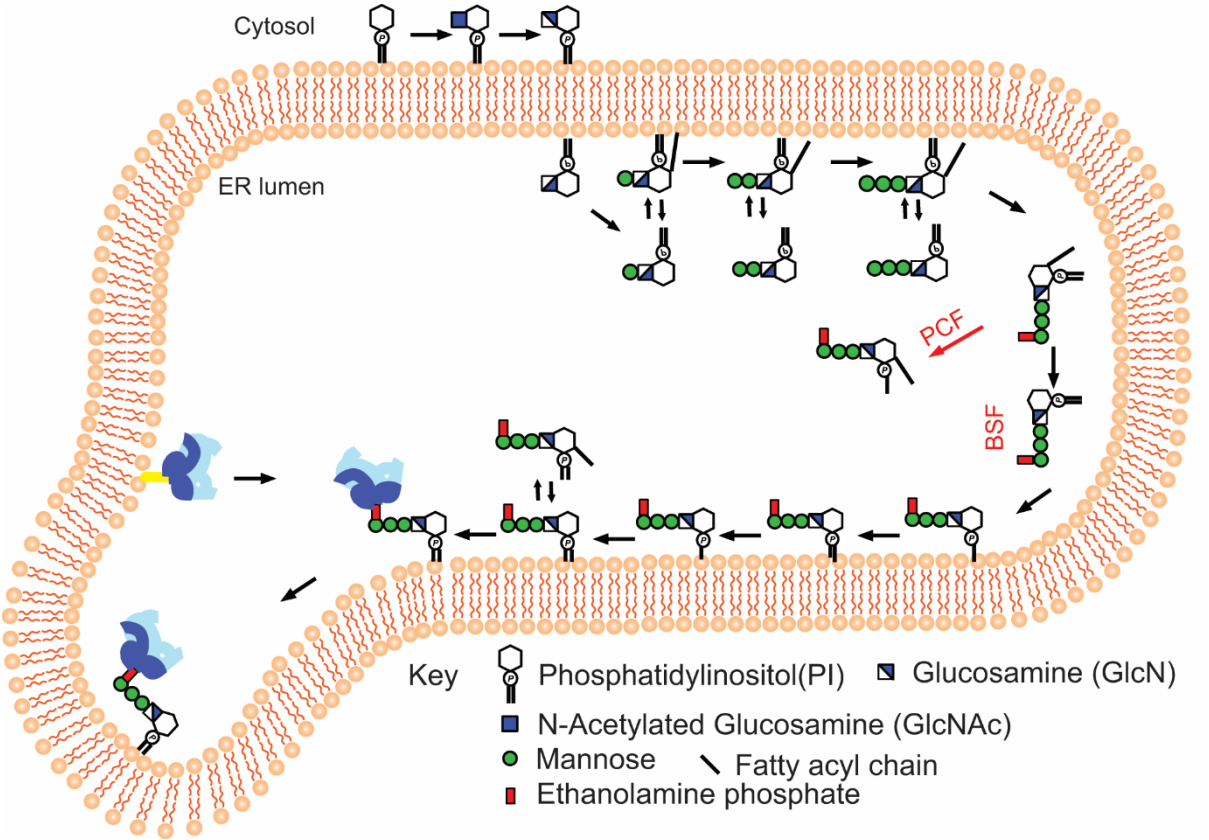

Figure 1.10 General scheme for GPI biosynthesis in the ER of *T. brucei* BSF.

The ER is depicted as a topologically closed compartment. The mature GPI precursor in PCF *T. brucei* is also indicated. Explanation of the pathway is in Section 1.3.2.

1.3.3 Comparison of GPI anchor biosynthesis pathways in *T. brucei*, mammalian cells and yeast

Despite conservation in the core structure of GPI anchors across eukaryotic evolution, notable differences between the *T. brucei* and mammalian GPI biosynthetic pathways occur from GlcN-PI onwards (Güther *et al*., 1994), including the addition of extra ethanolamine phosphate groups to mammalian anchors, fatty-acid remodelling of *T. brucei* GPI precursors and the timing of inositol acylation and deacylation (i.e., the addition of a fatty acid to the 2-OH of the *myo*-inositol residue). The latter occurs in *T. brucei* only after the addition of the first Man residue to GlcN-PI and intermediates from Man_1GlcN-(acyl)PI onwards are in equilibrium between their inositol acylated and non-acylated forms through the action of inositol acyltransferase(s) and inositol deacylase(s). In mammalian cells, inositol acylation occurs at the level of GlcN-PI and strictly precedes mannosylation to Man_1GlcN-(acyl)PI. Furthermore, the acyl chain remains attached once added until the mature GPI anchor is added to protein in the endoplasmic reticulum. Thereafter, it may remain but is usually removed by an inositol deacylase, prior to a form of lipid remodelling that is distinct from the fatty acid remodelling of *T. brucei* (Fujita and Kinoshita, 2012) (Kinoshita *et al*., 2013) (Kinoshita and Fujita, 2016). The lipid remodelling in yeast differs from both *T. brucei* and mammalian cells. It happens in the ER but after the GPI anchor precursor is transferred to protein. The unsaturated fatty acid at the sn-2 position is removed and replaced with a C26:0 chain. In some cases, the diacylglycerol is further exchanged for ceramide in the Golgi (Orlean and Menon, 2007).

1.4 Enzymes catalysing GPI anchor biosynthesis in *T. brucei*

1.4.1 Known enzyme complexes in the *T. brucei* GPI anchor biosynthesis pathway

Since the GPI anchor pathway is a potential drug target, significant work has been performed to characterise the enzymes of this pathway. However, it was the power of mammalian cell and yeast genetics that led to the identification of the majority of GPI biosynthesis genes, reviewed in (Kinoshita, 2020) (Pittet and Conzelmann, 2007). We currently have reasonably advanced models for GPI anchor biosynthesis and processing in trypanosomes, mammalian cells and yeast and the similarities and differences in these pathways have been reviewed extensively elsewhere (Kinoshita, 2020) (Kinoshita

and Fujita, 2016) (Jaquenoud *et al.*, 2008) (Pittet and Conzelmann, 2007) (Ferguson *et al.*, 2017). For most organisms, the functions and interactions of putative GPI pathway gene products have been inferred from experimental work in mammalian or yeast cells. In a few cases, these functions have been experimentally confirmed in *T. brucei*. Fig 1.11 is the GPI anchor biosynthesis pathway in *T. brucei* BSF and Table 1.1 is a list of enzymes (known or homologs of mammalian or yeast enzymes) that catalyse GPI anchor biosynthesis in *T. brucei*. The enzymes of this pathway have some common features:

(i) They tend to be very low abundance proteins. [The ranking of these proteins (from 1= most abundant, to 7150 = undetectable) in BSF *T. brucei* has been estimated by combining numerous deep proteomics datasets (Tinti *et al.*, 2019) (Ji *et al.*, 2021). The rankings (Table 1.1) show that several GPI pathway components are undetectable and that many (apart from several of the GPI transamidase complex) are in the bottom quartile of cellular protein abundance].

(ii) Where tested, they appear to be essential for the survival of BSF *T. brucei*. Experimental evidence has been reported for GlcNAc-PI de-N-acetylase (TbGPI12) (Chang *et al.*, 2002), the third mannosyltransferase (TbGPI10) (Nagamune *et al.*, 2000), the inositol deacylase 2 TbdeAc2 (Hong *et al.*, 2006) and TbGPI8 of the GPI transamidase (Lillico *et al.*, 2003)

(iii) They tend to function as protein complexes. Three known examples of GPI pathway complexes in *T. brucei* are the UDP-GlcNAc: PI α1-6 GlcNAc-transferase complex, the GPI transamidase complex (Lillico *et al.*, 2003) (Nagamune *et al.*, 2003) (Hong *et al.*, 2006) and the ethanolamine phosphate transferase 2 complex (Stokes *et al.*, 2014). The first step of GPI biosynthesis is the addition of GlcNAc to PI by a UDP-GlcNAc : PI α1–6 GlcNAc-transferase (GPI GnT) complex. The composition of this complex was determined in mammalian cells, where seven subunits have been identified: PIGA, PIGC, PIGH, PIGP, PIGQ and PIGY (a homologue of ERI1 first identified in yeast and shown to associate with GPI2 (Sobering *et al.*, 2004)), and DPM2 (Table 1.1) (Kinoshita, 2020). The DPM2 component is a non-catalytic subunit of dolichol phosphate mannose synthetase. The complex was realised through a series of elegant functional cloning and co-immunoprecipitation experiments using individually epitope-tagged bait and prey components. A similar multi-subunit complex has been proposed in yeast where homologues for all the subunits, except DPM2, have been identified (Table 1.1) (Pittet and Conzelmann, 2007). However, experimental evidence for physical associations between most of these yeast subunits is lacking. The

T. brucei homologous of these components were found by BLASTp search, and the experimental determination of this complex is illustrated later in the results chapter. The ethanolamine phosphate transferase 2 complex contains TbGPI13, the catalytic subunit, that is stabilised by TbGPI11, as described in (Stokes *et al.*, 2014). With respect to the GPI anchor transamidase, the catalytic (TbGPI8) subunit and the four other subunits (TbGAA1, TTA1 and TTA2 and TbGPI16) have all been shown to belong to the same complex (Nagamune *et al.*, 2003) with TbGP16 being disulphide linked to TbGPI8 and shown to form a complex (Hong *et al.*, 2006).

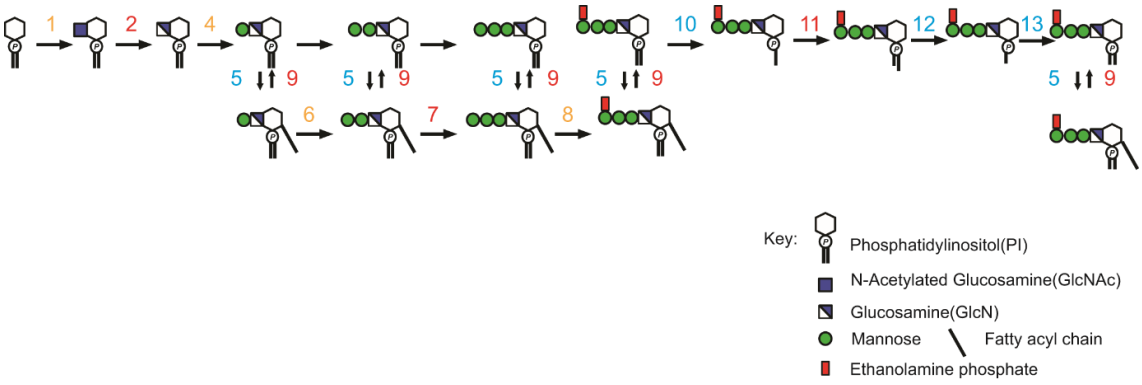

Figure 1.11 Scheme of GPI anchor biosynthesis pathway of BSF *T. brucei*.

Each step is annotated with numbers sequentially and the enzymes catalysing each step are listed in Table 1.1. Red annotated enzymes are experimentally determined known components in this pathway. Enzymes annotated in amber are homologous to known GPI pathway components in mammalian and/or yeast cells but without experimental data. Enzymes annotated in blue are the unknown "missing links" in the *T. brucei* GPI pathway.

Table 1.1 List of enzymes and protein components catalysing GPI anchor biosynthesis in *T. brucei*.

Catalysing steps	Enzymes	Genes	Protein ranking in BSF proteomics (Ji *et al.*, 2021)
1	GPI-GlcNAc transferase (GPI-GnT)	TbGPI3 (Tb927.2.1780) TbGPI15 (Tb927.5.3680) TbGPI19 (Tb927.10.10110) TbGPI2 (Tb927.10.6140) TbGPI1 (Tb927.3.4570) TbArv1 (Tb927.3.2480) TbERI1 (Tb927.4.780)	6475 7038 6185 6798 5285 7150 7150
2	GlcNAc-PI de-N-acetylase	TbGPI12 (Tb927.11.12080)	5818
3	Flippase	Not identified	
4	α1-4 mannosyltransferase I	TbGPI14 (Tb927.6.3300)	7150

5	Inositol acyltransferase	Not identified	
6	α1-4 mannosyltransferase Ⅱ	TbGPI18(Tb927.10.13160)	7150
7	α1-4 mannosyltransferase Ⅲ	TbGPI10 (Tb927.10.5560)	5212
8	EtNP transferase 2	TbGPI13 (Tb927.11.5070) TbGPI11 (Tb927.10.13290)	5700 2534
9	Inositol deacylation	TbdeAc (Tb927.10.4780) TbdeAc2 (Tb927.3.2610)	2693 5715
10	GPI-phospholipase A2	Not identified	
11	Lyso-GPI acyltransferase Ⅰ	TbGUP1(Tb927.10.15910)	4053
12	GPI-phospholipase A1	Not identified	
13	Lyso-GPI acyltransferase Ⅱ	Not identified	
14	GPI transamidase	TbGPI8 (Tb927.10.13860) TbGAA1 (Tb927.10.210) TbGPI16 (Tb927.4.1920) TTA1 (Tb927.11.15760) TTA2 (Tb927.10.5080)	1604 2503 1140 1557 7150

1.4.2 Other known GPI pathway enzyme complexes in mammalian cells and yeast

In addition to the GPI-GnT and GPI transamidase complexes, there are two more GPI pathway enzyme complexes known in mammalian cells and yeast.
GPI-mannosyltransferase I (GPI MTI) that in mammalian cells and yeast transfers the first mannose to GlcN-(acyl)PI and produce Man_1GlcN-(acyl)PI contains two subunits: The catalytic subunit PIG-M and its associated subunit PIG-X. PIG-M is an ER resident protein with its active site on the lumenal side of the ER membrane (Maeda *et al.*, 2001) and PIG-X binds to and stabilises PIG-M (Ashida *et al.*, 2005). The latter study also reported that Gpi14p and Pbn1p are the yeast homologous of PIG-M and PIG-X. The other GPI pathway complexes present in mammalian and yeast cells, but not in *T. brucei*, are the GPI ethanolamine phosphate transferase-2 (EtNP-T2) and ethanolamine phosphate transferase-3 (EtNP-T3) complexes. The common subunit in these two complexes PIG-F (Inoue *et al.*, 1993), stabilises both PIG-G (Shishioh *et al.*, 2005) and PIG-O (Hong *et al.*, 2000), which are the catalytic subunits for EtNP-T2 and EtNP-T3, respectively.

1.4.3 Enzymes involved in side-chain modifications of GPI-anchors

The GPI anchor side-chain modifications produce structural microheterogeneity and also differentiate different organisms. The BSF *T. brucei* GPI anchor side-chain consists

of galactose components (Ferguson *et al.,* 1988), whereas the PCF *T. brucei* GPI anchor side-chain carries sialylated poly-N-acetyllactosamine and poly-lacto-N-biose structures, reviewed in (Duncan *et al.,* 2021). However, there are only a few enzymes have been identified catalysing the steps of the side-chain modification. The trans-sialidase was the first enzyme found that modifies the GPI-anchored procyclins on the cell surface of PCF *T. brucei* (Engstler *et al*., 1993) (Pontes De Carvalho *et al.,* 1993) (Montagna *et al.,* 2002) (Montagna *et al*., 2006). Our lab has characterised three more enzymes that elaborate the GPI anchor side chain in PCF *T. brucei,* which are TbGT8, TbGT3 and TbGT10. TbGT8 is a UDP-GlcNAc : β-Gal-GPI β1-3 GlcNAc transferase (Izquierdo *et al.,* 2009) and TbGT10 is a UDP-GlcNAc : βGal β1-6 GlcNAc-transferase (Duncan *et al.,* 2021). The absence of their enzymatic activity does not impair the growth of the PCF *T. brucei* in culture but results in incomplete glycan branches of the GPI-anchor. TbGT3 is a UDP-Gal : β-GlcNAc-GPI β1-3 Gal transferase. The PCF conditional null mutant of *TbGT3* also shows a reduction in the molecular weight of the GPI-anchor side-chain glycans, but no defect in either cell growth or ability to infect tsetse flies (Izquierdo *et al.,* 2015).

In mammalian cells, βGalNAc appears to be a common side-chain modification of GPI anchors, catalysed by a GPI-specific GalNAc transferase PGAP4 (Hirata *et al.,* 2018). Later, through CRISPR-Cas9 mediated phenotypic screening, Kinoshita's group showed that the β1,3-galactosyltransferase 4 (B3GALT4), previously characterized GM1 ganglioside synthase, has a dual function in transferring galactose to a β1,4-linked GalNAc side-chain of GPI (Wang *et al.*, 2020).

2 Aims

The aim of this project was to take new approaches to try to identify currently unidentified proteins that catalyse, or participate in, key steps of GPI anchor biosynthesis in *T. brucei* (Fig 2.1).

The basic idea was to epitope tag known components of the GPI biosynthetic pathway and perform affinity immunoprecipitation experiments with antibodies to the tags. In parallel, I performed affinity immunoprecipitation from non-tagged cells to exclude non-specifically bound proteins. We then investigated label-free, cross-linking and stable isotope labelling in cell culture (SILAC) proteomics methodologies (Fig 2.1). We hope that some of the proteins specifically pulled down with the tagged GPI pathway enzymes may also be GPI pathway components, based on the fact that it is quite common for proteins involved in biosynthetic pathways to be physically associated with each other.

We tested this idea by seeing whether the pull-down of the GlcNAc-PI de-N-acetylase (TbGPI12) also pulls-down the first mannosyltransferase (MTI), and vice-versa. This association has been predicted from biochemical data that suggest substrate channelling between the two enzymes (Smith *et al.*, 1996). Another test was to pull down the main catalytic subunit of GPI GlcNAc transferase, TbGPI3, to see if other subunits that are predicted to be associated with it can also be pulled down. We then applied the approaches and conditions successful to either of these test situations to the pull-downs of the tagged GPI enzymes such as TbdeAc2 (Hong*et al.*, 2006) and TbGup1 (Jaquenoud *et al.*, 2008), that catalyse inositol-deacylation and the re-acylation the GPI anchor glycerolipid moiety (via acyl- (myristoyl-) transferase activity), respectively. With respect to the TbdeAc2-tag and TbGup1-tag pulldowns, our hypothesis was that we may also capture upstream and downstream 'missing components' of the fatty acid remodelling machinery. For example, we might expect GPI-phospholipase A2 (PLA2), GPI-phospholipase A1 (PLA1), GPI-inositol-acyltransferase and/or fatty acid remodelling myristoyltransferases to be found in this way (see Fig 1.11). We would look for sequence motifs typical of such enzymes (for example, GDSL/A lipase or HDXXY and YXGXG PLA2 motifs) in the co-immunoprecipitated proteins. If so identified, we would attempt to construct conditional null mutants of these genes to assess (a) whether they are essential to the parasite *in vitro* and/or *in vivo*, and therefore

potential drug targets, and (b) to prepare cell-free systems from parasites where the proteins are missing (or at very low levels just prior to cell death) to provide biochemical evidence for the activity of the gene products. Tagged-proteins may also be immunoprecipitated to perform on-bead digestions to test the enzymatic activity using appropriate substrates.

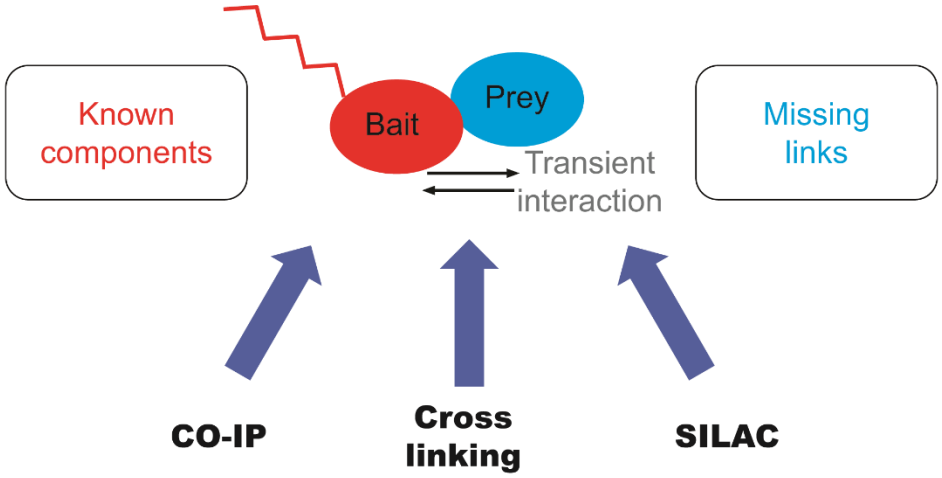

Figure 2.1 Overview of approaches to be applied to fulfil the aims of the project.

3 Materials and Methods

3.1 Reagents

All general chemicals were purchased from Sigma or VWR BDH unless otherwise indicated. HPLC grade solvents were obtained from VWR BDH with the exceptions of acetonitrile (Thermo) and methanol (Sigma).

3.2 *Trypanosoma brucei* cell culture

3.2.1 *T. brucei* bloodstream form (BSF) cell lines and cell cultures

T. brucei brucei Lister strain 427 bloodstream form (BSF) parasites expressing VSG variant 221 (also known as MITat1.2) and transformed to stably express T7 polymerase and the tetracycline repressor protein (TetR) under G418 antibiotic selection (Wirtz *et al.*, 1999) was used in this study and will be referred as bloodstream form wild type (BSF WT). BSF WT cells were cultivated in HMI-11T medium containing 2.5 µg/mL G418 at 37˚C in a 5% CO_2 incubator as previously described (Wirtz *et al.*, 1999). HMI-11T is a modification of the original HMI-9 (Hirumi and Hirumi, 1989) that uses 56 mM 1-thioglycerol in place of 200 mM 2-mercaptoethanol, and contains 10% heat inactivated fetal bovine serum (PAA) and lacks of serum plus (Hazleton Biologics, Lenexa, Kansas) (Table 3.1).

Bloodstream form *T. brucei* Lister 427 (MITat 1.2), clone 221a, 2T1 cells (Alsford *et al.*, 2005) were used for the RNA interference (RNAi) knock down using the pRPa construct in this study and will be referred as bloodstream 2T1 wild type (BSF 2T1 WT). BSF 2T1 WT cells were cultivated in HMI-11T medium containing 1 µg/mL phleomycin, 1 µg/mL puromycin.

Bloodstream form *T.brucei* Lister 427 (MITat1.2), clone 221a, $2T1^{T7\text{-}Cas9}$ cells (Rico *et al.*, 2018) were used for CRISPR Cas9 mediated gene manipulations and will be referred as bloodstream form Cas9 wild type (Cas9 WT). Cas9 WT cells were cultivated in HMI-11T medium containing 2 µg/mL blasticidin and 1 µg/mL hygromycin.

All BSF cells were maintained in culture to mid log (1-2 × 10^6 cells/mL) and then sub-cultured to fresh medium at a density of 1-2 × 10^4 cells/mL.

3.2.2 *T. brucei* procyclic form cell line and cell culture

T. brucei brucei Lister 427 procyclic form (clone 29.13.6) that stably expresses T7 polymerase and tetracycline repressor (TetR) protein under G418 and hygromycin antibiotic selections, respectively, was used in this study and will be referred as procyclic wild type (PCF WT). PCF WT was cultivated in SDM-79 medium (Table 3.1) containing 15 µg/mL G418 and 50 µg/mL hygromycin in a 28 °C incubator to the mid log (3-5 × 10^7 cells/mL) and sub-cultured to fresh medium at a density of 1-2 × 10^6 cells/mL.

Table 3.2 lists the concentration of antibiotics used to maintain selection for the cell lines used in this study.

Table 3.1 Full composition of culture media.

Medium	Life stage	pH	Composition (for 10L)	Supplier
HMI-11T	*T. brucei* bloodstream form	7.4	18 g HMI-9 powder	Invitrogen
			36 mM $NaHCO_3$	VWR
			2 mM glutamine	Invitrogen
			10% FBS heat inactivated	Thermo Fisher
			57.5 nM 1-Thioglycerol	Sigma
SDM-79	*T. brucei* procyclic form	7.3	25 g SDM-79 powder	Invitrogen
			24 mM $NaHCO_3$	VWR
			11.5 µM haemin	Sigma
			0.2 mM glutamine	Invitrogen
			10% FBS heat inactivated	Thermo Fisher

Table 3.2 Antibiotics used in the cell culture and genetic manipulation of *T. brucei*.

Antibiotic	Resistance gene	Concentration (µg/mL)			Supplier
		BSF WT	BSF 2T1 WT/ Cas9 WT	PCF WT	
Neomycin (G418)	Neomycin phosphotransferase (NEO)	4		15	Gibco
Hygromycin	Hygromycin phosphotransferase (HYG)	4	1	50	Roche
Puromycin	Puromycin acetyltransferase (PAC)	0.1	1	1	Sigma
Phleomycin	Phleomycin resistance protein (PHLEO)	1	1	1	Invitrogen
Blasticidin	Blasticidin S deaminase	2.5	2	10	Invitrogen
Tetracycline		0.5	1	0.5	Calbiochem

3.3 General *T. brucei* cell culture protocols

3.3.1 Cell density determination

To determine the cell density, *T. brucei* bloodstream form cells were counted on a Neubauer haemocytometer directly or diluted 1:10 in 1× trypanosome dilution buffer (TDB) (Table 3.3). While for *T. brucei* procyclic form cells were first diluted 1:10 in counted 1× phosphate buffer saline (PBS) (Sigma) or in SDM-79 medium with viability assessed and cells counted on a haemocytometer.

Table 3.3 5 × Trypanosoma Dilution Buffer (TDB).

Composition	pH	Chemicals/L (g)
KCl		1.86
NaCl		23.4
MgSO4·7H2O	7.4	1.23
Na_2HPO_4		14.2
$NaH_2PO_4·2H_2O$		1.56
Glucose		18

3.3.2 Generation of stabilates for cryo-preservation

For generation of the stabilates for cryo-preservation, 10 mL of *T. brucei* bloodstream form cells were grown to mid log phase (1-2 × 10^6 cells/ml) and collected by centrifugation (800 g for 10 min at 4°C). The cell pellets were resuspended in 1.5 mL HMI-11T medium containing 10% glycerol without antibiotics. Aliquots of 0.5 mL were transferred to 2 mL Cryovials (Alphalab), frozen overnight at -80 °C and then transferred to liquid nitrogen for long-term storage.

For *T. brucei* procyclic form, 10 mL cells were grown to mid-log phase (2 × 10^7 cells/ml) and centrifuged (600 g for 10 min at 4°C), the pellets were resuspended in 1.5 mL SDM-79 containing 10% glycerol without antibiotics. Aliquots of 0.5 ml were transferred to Cryovials, frozen overnight at -80 °C and then transferred and stored under liquid nitrogen.

3.3.3 Generation of clonal cell lines

For bloodstream *T. brucei* cell lines, clonal parasite cell lines were obtained by limit dilution. Mid-log parasites were diluted to a density of 1.65 cells/mL in fresh culture medium with suitable antibiotics and then dispensed into a 96 well plate with 200 µl/well. Growth was monitored over the following 5-7 days and cells collected for further analysis.

3.3.4 *T. brucei* bloodstream form electroporation

Bloodstream form cells were grown to mid log phase density (1-2 × 10^6 cells/mL), centrifuged (800 g for 10 min at 20 °C) and resuspended in either 100 µL Human T-Cell Nucleofector® solution containing supplement (Lonza) or 100 µL cytomix buffer (Table 3.4) made in house, transferred into the supplied cuvette and subjected to electroporation using Amaxa Nucleofector™ electroporator with program X-001, or program Z-001 for CRISPR Cas9 edition. For each cuvette, 1 × 10^7 cells alongside 1–20 µg of ethanol-precipitated DNA (Section 3.4.15) were used for each transfection. One cuvette without DNA was served as a negative control for each experiment. After the electroporation, the cells from each cuvette were transferred to 10 mL or 50 mL HMI-11T medium and allowed to recover for 6-8 h or overnight at 37°C before adding selection antibiotics. The choice of DNA amount and cell recovery time ranges depended on the experimental purposes. For *in situ* epitope tagging, 20 µg DNA was used per transfection and the cells were recovered overnight before adding selection antibiotics. For generation overexpression or RNAi knock down cell lines, 2 µg linearised plasmid was used and the cells were recovered for 6-8 h before antibiotics were added. For knock out constructs, 10 µg DNA was used for transfection and the recovery time for the cell line was overnight. For CRISPR Cas9 mediated gene manipulation transfection, single guide RNA (sgRNA) and repair templates mixture (20 µg) was electroporated to the cells and the antibiotic was added after the cells recovered overnight.

Table 3.4 Cytomix buffer.

Composition	pH
2 mM EGTA	
120 mM KCl	
25 mM HEPES	

5 mM $MgCl_2 \cdot 6H2O$	
0.5% Glucose	7.6
100 µg/mL defatted BSA	
1.0 mM Hypoxanthine	
10 mM KH_2PO_4	
10 mM K_2HPO_4	

3.4 Molecular biology protocols

3.4.1 *In silico* analyses and searches of DNA and protein sequences

For *T. brucei*, DNA and amino acid sequences for genes of interest were retrieved from TriTrypDB (Aslett *et al.*, 2010) in FASTA format (http://tritrypdb.org/tritrypdb/). Untranslated regions of genes were identified based on the latest released whole genome sequencing (Müller *et al.*, 2018) for *T. brucei* Lister 427 strain but also taking the annotations for *T. brucei* 927 strain into account.

For identification of putative subunits of UDP-GlcNAc : PI α1-6 GlcNAc-transferase (GPI GnT) complex in *T. brucei*, conventional BLASTp with default settings (Acland *et al.*, 2012) or Domain Enhanced Lookup Time Accelerated BLASTp (Boratyn *et al.*, 2012) were applied to find the homologous. The protein sequences of mammalian cells and yeast were retrieved from UniProt (Bateman *et al.*, 2021), and the gene IDs of the query input sequences are listed in (Table 5.1). Protein sequence multiple alignments were assembled and visualised using ClustalW (Chenna *et al.*, 2003). *In silico* plasmid constructs sequences used in this study were all assembled using CLC Main Benchwork (CLC Bio). The applications in this software also allowed for the design of oligonucleotide primers, sequence alignment, and *in silico* cloning and endonuclease digestion mapping.

3.4.2 Primers

Primers used for different applications in this study were listed from (Table 3.5) to (Table 3.11). Primers were designed using CLC Main Benchwork with following features: i, they should contain ~45 – 55 % guanine-cytosine (GC) content and the 3' end of the primer should have a guanine (G) or cytosine (C); ii, the melting temperature (T_m) of the primers were mainly calculated by the NEB T_m Calculator (https://tmcalculator.neb.com/#!/main) and less than 5 °C difference in T_m between primer pairs were designed. The primers were synthesised either by Thermo Fisher or the Oligonucleotide Synthesis Service, University of Dundee.

3.4.3 Polymerase chain reactions (PCR)

Polymerase chain reactions (PCR) for DNA amplification were carried out using Q5® High-Fidelity DNA polymerase (New England Biolabs) or LongAmp® DNA polymerase according to the manufacturer's instructions. The template concentrations used ranged from 1 – 10 ng for genomic DNA (gDNA), 10 – 100 pg for purified plasmid DNA and optimised for each target of interest. The general conditions and programs used for each polymerase are listed in (Table 3.12).

Table 3.5 List of primers used for *in situ* epitope tagging proteins in this study.

Oligo Names	Description	Gene ID	Sequence	Lower case letters
OL1 (Forward)	*In situ* epitope tagging TbGPI3 with 3 c × Myc	Tb927.2.1780	5′-TGATTGATATTGCACCAGATTTTCCACTGGAGTTGTACTCTCGTAACCGGGAGAAGCTTCAAGTTGTGGGAAGCCCATCCgaacaaaagctgggtacc-3′	Specific annealing to pMOTag43M vector
OL2 (Reverse)			5’-CAACGCGAAACAATGACAGAGAGAGAGAGAGAAGGGCGAAAACAAAAAGGAT CGCGGTAGAGAGGACCCCGCCCATACCCctattcctttgccctcggac-3’	Specific annealing to pMOTag43M vector
OL3 (Forward)	*In situ* epitope tagging TbdeAc2 with 3 × c-Myc	Tb927.3.2610	5’-CCTTTCCATGTTTTTTGCTTGGTCCGGCGCTCGTGTCCGGTGGTAGTGGTGGCACCCTTCTAGCTCCTGGTACCTCAAACgaacaaaagctgggtacc-3’	Specific annealing to pMOTag43M vector
OL4 (Reverse)			5’-GAAAAAAAACAAAAAGGAAAGGAAAGGAACACGAGAGCCGTGATTCCAAAAGCAAAGTTAATATACACGCTTTTATTTTCctattcctttgccctcggac-3’	Specific annealing to pMOTag43M vector
OL5 (Forward)	*In situ* epitope tagging Tb927.11.11740 with 3 HA	Tb927.11.11740	5’-GTGTCATCTTGTCAATTGGCATCTTGCACTGGTGCCCACTTTTGTGTCTTGGTGTATCGCCGAAAGATCCCACCGTAAGGgaacaaaagctgggtacc-3’	Specific annealing to pMOTag4H vector
OL6 (Reverse)			5’-CCTTTGTTGACTCTTAAGAAAAAAAAACCCATAATACCTTTGAGAACCCTCTGCCTCCCTTTCATAACTGTGGTCGCTGCctattcctttgccctcggac-3’	Specific annealing to pMOTag4H vector
OL 7 (Forward)	*In situ* epitope tagging TbdeAc2 with 12 × c-Myc	Tb927.3.2610	5’-cccaagcttGCACACGGCTGTTGTTATG-3’	*Hind*III
OL 8 (Reverse)			5’-tgctctagaGTTTGAGGTACCAGGAGCTA-3’	*Xba*I

Table 3.6 List of primers for generation the RNAi knock down and overexpression plasmid.

Oligo Names	Description	Gene ID	Sequence	Restriction sites
OL9 (Forward)	Generation of RNAi knock down Tb927.11.1140 & Tb927.11.11750 with p2T7 plasmid	Tb927.11.1140 & Tb927.11.11750	5′-CCG*CTCGAG*TGTCCAGGTGGTTCATCGTG-3′	*Xho*I
OL10 (Reverse)			5’-CG*GGATCC*TCCCTCCCCATCGTACCAATA-3’	*BamH*I
OL11 (Forward)	Generation of RNAi knock down Tb927.11.1140 & Tb927.11.11750 with pRPa^{iSL} plasmid	Tb927.11.1140 & Tb927.11.11750	5’-CGCG*GGTACC*GGGCCCTGTCCAGGTGGTTCATCGTG-3’	*Kpn*I/*Apa*I
OL12 (Reverse)			5’-CGCG*GGATCC*TCTAGATCCCTCCCCATCGTACCAATA-3’	*BamH*I/*Xba*I
OL13 (Forward)	Generation of plew100_v1 Tb927.11.11750 plasmid	Tb927.11.11750	5’-ACCAAAAAGTAAAATTCAC*AAGCTT*ATGATGCGTCCACTCTCAC -3’	*Hind*III
OL14 (Reverse)			5’-AAAGCCAACTAAATGGGCA*GGATCC*TTAGACAGCCATCTTTTGTTTAAAC-3’	*BamH*I

Table 3.7 List of primers for Gibson assembly in this study.

Oligo Names	Description	Gene ID	Sequence
OL15(Forward)	Amplification of CDS of Tb927.11.11750 for Gibson assembly	Tb927.11.11750	5'-GTAAAATTCACAAGCTTTAGATGATGCGTCCACTCTCAC-3'
OL16(Reverse)			5'-CCAACTAAATGGGCAGGATCCTTAGACAGCCATCTTTTGTTTAAAC-3'
OL17(Forward)	Amplification of linearised plew100_v5 vector for Gibson assembly		5'-GGATCCTGCCCATTTAGTTG-3'
OL18(Reverse)			5'-CTAAAGCTTGTGAATTTTACTTTTTG-3'
OL19 (Forward)	Amplification of 12 × c-Myc with optimised codon for Gibson assembly	Tb927.11.11750	5'-ATATCGGATCCGGCCCCCGCTCGAGGAGCAAAAGCTTATTAGCGAG-3'
OL20 (Reverse)			5'-TACGCCAAGCTTTGGCACGCGTCGACGCGCGGGAGATCCTCCTCCGA-3'
OL21 (Forward)	Amplification of linearised plew100_v5 Tb927.11.11750 plasmid for Gibson assembly		5'-TCGGAGGAGGATCTCCCGCGCTACGTGCTGGAGCTTGTCCA-3'
OL22 (Reverse)			5'-CTCGCTAATAAGCTTTTGCTCCGAGCAATTTGCGCCCGTAG -3'
OL23 (Forward)	Amplification of 5'UTR of Tb927.11.11740 for Gibson assembly	Tb927.11.11740 & Tb927.11.1150	5'-ACCATGATTACGCCAAGCTTGCATGGAACGAACGATGTCCATCTCA-3'
OL24 (Reverse)			5'-GTACTCGGTCATAATGCGTTACGGCTTGTGTG-3'
OL25 (Forward)			5'-GCCGTAACGCATTATGACCGAGTACAAGCCCACGGTG -3'

OL26 (Reverse)	Amplification of PAC for Gibson assembly		5’-ACCTGTATAACATctaGGCACCGGGCTTGCG -3’
OL27(Forward)	Amplification of 3’UTR of Tb927.11.11750 for Gibson assembly		5’-CCCGGTGCCTAGATGTTATACAGGTCGCTATG -3’
OL28(Reverse)			5’- GGTACCCGGGGATCCTCTAGAGTCGAGAGCCTGAGCTTTGAAGGTA-3’
OL29(Forward)	Amplification of linearised vector pUC57		5’- CGACTCTAGAGGATCCCCG-3’
OL30(Reverse)			5’-CATGCAAGCTTGGCGTAATC-3’
OL31(Forward)	Amplification of HYG for Gibson assembly	Tb927.11.1150	5’-AAACCCCGGCATTATATCGAAGCTTATGAAAAAGCCTGAACTCAC-3’
OL32(Reverse)			5’-ATAGCGACCTGTATAACATGGATCCCTATTCCTTTGCCCTCGG-3’
OL33(Forward)	Amplification of linearised pUC57-PAC KO construct for Tb927.11.11750		5’-GGATCCATGTTATACAGGTCGCTATG-3’
OL34(Reverse)			5’-AAGCTTCGATATAATGCCGGGGTTTA-3’

Table 3.8 List of primers for real time RT-qPCR.

Oligo Names	Description	Gene ID	Sequence
OL35 (Forward)	Amplicons of identical region of Tb927.11.11740 & Tb927.11.11750	Tb927.11.1140 & Tb927.11.11750	5′- ACGACTGGATGCTGTTGCAACTG-3′
OL36 (Reverse)			5’-ATTCCCAAAAGAAGTTCCTGTGCCAA-3’
OL37 (Forward)	Amplicon of specific region of Tb927.11.11750	Tb927.11.11750	5’-ATGCGGATGAAGCGATTCTCACA-3’
OL38 (Reverse)			5’-TCACCGGGTACATTTCAACCATG-3’

Table 3.3.9 List of primers for sgRNA used for CRISPR Cas9 mediated gene manipulation

Oligo Names	Description	Gene ID	Sequence
OL39 (Forward)		Tb927.11.1140 & Tb927.11.11750	5′- TAATACGACTCACTATAGGGAACTATGCAGCCCCGACA

	Amplification of sgRNA for CRISPR edition for Tb927.11.11740 & Tb927.11.11750	GCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GC-3′
OL40 (Reverse)		5’- GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGAC TAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3’

Table 3.10 List of primers used for probe synthesis for Southern blot.

Oligo Names	Description	Gene ID	Sequence
OL41 (Forward)	Amplification of 1000 bp identical region of Tb927.11.11740 & Tb927.11.11750	Tb927.11.1140 & Tb927.11.11750	5′- ATGATGCGTCCACTCTCACTTG-3′
OL42 (Reverse)			5’-ACACTTCCGTAGTGTCGTCA-3’
OL43 (Forward)	Amplification of PAC ORF		5′- ATGACCGAGTACAAGCCCACG-3′
OL44 (Reverse)			5′- TCAGGCACCGGGCTTGCG-3′
OL45 (Forward)	Amplification of HYG ORF		5′- ATGAAAAAGCCTGAACTCAC-3′
OL46 (Reverse)			5′- CTATTCCTTTGCCCTCGGA-3′

Table 3.11 List of primers for genotype check for mutant cell lines.

Oligo Names	Description	Gene ID	Sequence
OL47 (Forward)	Amplification of CRISPR Cas9 mutant clone 1 gene rearrangements region. (Section 6.6.3)	Tb927.11.1140 & Tb927.11.11750	5′- ATAACACTGTGTAGTGCGA-3′
OL48 (Reverse)			5’-TTAGACAGCCATCTTTTGTTTAAAC-3’
OL47 (Forward)	Amplification of original locus of *Tb927.11.11750* (section 6.8.2)	Tb927.11.11750	5′- ATAACACTGTGTAGTGCGA-3′
OL49 (Reverse)			5′- TCAGGCACCGGGCTTGCG-3′
OL51 (Forward)			5′- ATGAAAAAGCCTGAACTCAC-3′

OL52 (Reverse)	Amplification of BSD replaced *Tb927.11.11750* locus (section 6.8.2)	5′- TTAGCCCTCCCACACATAAC-3′

Table 3.12 PCR amplification reactions and program.

PCR reaction of Q5® High-Fidelity DNA polymerase			
Component		**Volume (µL)**	**Final concentration**
5× Q5 reaction buffer		10	1×
10 mM dNTP mix		1	200 µM
10 µM forward primer		2.5	0.5 µM
10 µM reverse primer		2.5	0.5 µM
Template DNA		variable	< 1,000 ng
Q5 High-Fidelity DNA Polymerase		0.5	1 U
5× Q5 High GC Enhancer		5	1 ×
ddH_2O*		to 50 µl	
Thermo cycle programme			
Step	**Temperature (°C)**	**Time (sec)**	**Number of cycles**
Initial denaturation	98	30	1
Denaturation	98	10	35
Annealing	50 - 72	30	
Elongation	72	20 - 30 /kb	
Final extension	72	120	1

PCR reaction of LongAmp® *Tag* DNA polymerase			
Component		**Volume (µl)**	**Final concentration**
5× LongAmp Taq Reaction Buffer		10	1×
10 mM dNTP mix		1.5	300 µM
10 µM forward primer		2	0.4 µM
10 µM reverse primer		2	0.4 µM
Template DNA		variable	< 1,000 ng
LongAmp Taq DNA Polymerase		2	5 U
ddH_2O		to 50 µl	
Thermo cycle programme			
Step	**Temperature (°C)**	**Time (sec)**	**Number of cycles**
Initial denaturation	94	30	1
Denaturation	94	30	30
Annealing	45 - 65	50	
Elongation	65	20 - 30 /kb	
Final extension	65	600	1

*: ddH_2O = double-distilled water

3.4.4 Agarose gel electrophoresis

DNA samples in this study were analysed by agarose gel electrophoresis. Agarose (VWR) were prepared in 1 × TAE buffer (40 mM Tris, 20 mM acetic acid, and 1 mM

EDTA) at 0.8% (w/v) containing 0.4 µg/mL ethidium bromide. For loading, DNA samples were mixed with 1× Blue Loading Dye (Promega) while 1kb DNA ladders (Promega) were served by side to indicate the size of the DNA fragments. The electrophoresis was carried out at 40-120 V in 1× TAE buffer with BioRad Mini-Sub Cell GT tanks until the desired separations for DNA fragments were achieved. The gels were visualised and imaged with a GelVue UV Transilluminator on a UGenius gel documentation system (Syngene).

3.4.5 Gel extraction and PCR purification of DNA fragments

Plasmid vectors or DNA fragments digested with endonuclease enzymes were separated and visualised using a UV transilluminator (Merck), gel bands of interest were excised, and DNA fragments were extracted with a Gel Extraction kit (Qiagen) according to manufacturer's instructions. PCR amplified DNA fragments were cleaned up using a PCR Purification kit (Qiagen) following the manufacturer's instructions. Final purified DNA was eluted with 2 mM TrisHCl (pH 8.5).

3.4.6 Quantification of DNA concentration of purity

The concentration of DNA was quantified on a NanoDrop 1000 Spectrophotometer (NanoDrop) through the absorption at 260 nm (A260), the purity of the DNA was estimated by checking the ratio of A260 to A280 (nucleic acid to protein).

3.4.7 Restriction endonuclease digestions

All endonucleases that were used in this study were purchased from New England Biolabs (NEB) and used according to the manufacturer's instructions. Analytical digestions were carried out for 1-2 h while preparative digestions were performed overnight. For double digestions, the optimum reaction conditions were checked using the Double Digest Finder tool (https://nebcloner.neb.com/#!/redigest). Digestion mxi or an aliquot were then analysed by agarose gel electrophoresis (Section 3.4.5).

3.4.8 Transformation of *E. coli* competent cells

For routine transformations of the ligation reactions, DH5α competent cells from MRC PPU reagents units (University of Dundee) were used. 50 µL competent cells were mixed with 1-5 µL 1-100 ng purified plasmid DNA or the ligation reaction generated with T4 ligase from Promega (Section). Cells were incubated on ice for 20 minutes before heat-shocked at 42 °C for 30 seconds, following by cooling on ice for 2 minutes. 1 mL Super Optimised Broth (SOC) medium were added to the cells and the transformed cells were incubated at 37 °C with agitation at 200 rpm for 1h. 50-200 µL aliquot of the cells was then spread onto Lysogeny Broth (LB) agar plate containing appropriate antibiotic for selection. Plates were then incubated overnight at 37 °C and individual colonies were then selected for further analysis. For Gibson assembly reactions, NEB® 5α Competent *E. coli* cells (NEB) within the kit was used for the transformation while for the T/A cloning for sequencing the Cas9 mutant clone 1 (Section 6.6.3) the competent cells that was supplied within the kit (Aligent, StrataClone PCR Cloning Kit) was used. The transformations for these two experiments were performed following the manufacturer's instructions.

3.4.9 Bacterial culture and storage

Recovered colonies from the transformation were picked up and inoculated into LB liquid medium with suitable antibiotic (Ampicillin 50 µg/mL), the bacteria were cultured overnight at 37°C with agitation at 200 rpm and subjected for plasmid DNA extraction. For long term storage, bacteria stabilates were prepared by mixing equal volumes of bacteria culture with 80% (v/v) glycerol. The stabilates were flash-frozen on dry-ice prior transferring to -80°C freezer. All media were prepared and sterilised by autoclavation by Central Technical Services at the University of Dundee.

3.4.10 Purification of plasmid DNA

Plasmid DNA was purified from *E. coli* cell pellet harvested from 5 mL LB culture using the QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. For large scale preparation, 100 mL bacterial cells were prepared at log phase and send for Maxiprep by Sequencing Service, University of Dundee.

3.4.11 Dephosphorylation

To minimise the self-ligation background when performing the ligation, plasmids that were digested with endonuclease were dephosphorylated prior to the ligation reaction. The plasmid was incubated with Antarctic Phosphatase (NEB) according to manufacturer's instructions and the reactions was stopped by heat-inactivation at 80 °C for 20 minutes.

3.4.12 Ligation

Linearised plasmids and endonuclease digested inserts that contained cohesive ends were ligated using T4 DNA ligase (NEB or Promega). Plasmid backbone and insert DNA fragments were mixed in equimolar with a ratio of 1:1 or 1:3 in 10 µL reaction volume with supplement of 1× Ligase Buffer and T4 DNA ligase. In general, 100 ng of the plasmid DNA was used for the reaction and an additional ligation without insert was performed as an indication of self-ligation of the plasmid. Reactions were incubated for 1-2 h at room temperature or overnight at 4 °C followed by transformation described in Section 3.4.8.

For T/A end cloning, the ligation was set up following the instructions of StrataClone PCR Cloning Kit (Agilent) and the transformation was described in Section 3.4.8.

3.4.13 Gibson Assembly

For generation the plew100_v5 internal 12 × c-Myc tagged Tb927.11.11750 plasmid and the knock out constructs for Tb927.11.11740 and/or Tb927.11.11750, Gibson Assembly reactions were performed to DNA fragments following manufacturer's instruction of NEBuilder® HiFi DNA Assembly kit(Table 3.13) (NEB). The reactions were then transformed to supplied competent bacteria cells according to the manufacturer's instruction.

Table 3.13 Gibson assembly reaction.

Component	2-3 Fragment Assembly	4-6 Fragment Assembly
DNA molar ratio	vector:insert = 1:2	vector:insert =1:1
Total amount of fragments	0.03-0.2 pmol (X µL)	0.2-0.5 pmol (X µL)
NEBuilder HiFi DNA Assembly Master Mix	10 µL	10 µL
ddH_2O	(10 – X) µL	(10 – X) µL

Total volume	20 µL	20 µL

3.4.14 DNA sequencing

DNA sequencing in this study was performed by the DNA Sequencing and Services (MRC PPU, University of Dundee, http://www.dnaseq.co.uk.) using Eurofins MWG Operon (i54 Business Park, Valiant Way, Wolverhampton, WV9 5GB) SmartSeq kit sequencing by cycle sequencing technology on Applied Biosystems model 3730XL DNA sequencer. DNA sequencing data was the analysed on CLC Main Workbench (CLC bio).

3.4.15 Ethanol precipitation

To concentrate and purify DNA for transfection into *T. brucei*, a 10% volume of 3 molar (M) sodium acetate pH 5.3 and 3 × volume of 100 % ethanol were added to the DNA sample and the mixture was gently mixed and incubated at -20 ℃ for overnight. Precipitated DNA was centrifuged at 13,000 × g for 10 min at 4 ℃ and washed three times with 70% ice cold ethanol, the supernatant was removed by centrifugation at 13,000 × g for 10 min at 4 °C. Resulting DNA pellet was air-dried in sterile hood and re-suspended with sterile water for further usage.

3.4.16 Quantitative reverse transcription polymerase chain reaction (RT-qPCR)

To perform quantitative reverse transcription polymerase chain reaction (RT-qPCR) on QuantStudio™ 3 Real-Time PCR System for checking the transcriptional level of gene of interest, total RNA was extracted from 1×10^7 *T.brucei* cells using RNAeasy kit (Qiagen) according to manufacturer's instructions. RT-qPCR reaction was then set up for RNA samples using the Luna Universal One-Step RT-qPCR Kit (NEB). The protocol used for the RT-qPCR Kit was summarised in Table 3.14.

Table 3.14 RT-qPCR reaction and program.

RT-qPCR Reaction mix		
Component	Volume (µL)	Final quantity
Luna Universal One-Step reaction Mix (2×)	10	1 ×
Luna WarmStart® RT enzyme mix (20×)	1	1 ×
10 µM forward primer	0.8	0.4 µM

10 µM Reverse primer		0.8	0.4 µM
20 ng µl^{-1} template RNA		5	100 ng
Nuclease-free water		to 20	
Thermal cycles			
Cycle Step	Temperature (°C)	Time (sec)	Number of cycles
Reverse transcription	55	600	1
Initial denaturation	95	10	1
Denaturation	95	10	40
Extension	60	30	
Melt curve	95	15	1
	60	60	
	95	15	

3.4.17 Southern blotting

3.4.17.1 Probe synthesis

The DNA probes used in this study were digoxigenin (DIG)-labelled. These were generated using PCR DIG Probe Synthesis Kit (Roche) according to the protocol summarised in Table 3.15. For each probe, a reaction which uses normal dNTP solution was included as a control to indicate the specificity and efficiency of the DIG-labelled probe synthesis. Pair of primers for synthesis different probes were listed in (Table 3.10).

3.4.17.2 DIG-labelled probe detection

For analysis of *T. brucei* mutants, 5 µg of gDNA from the clones were digested solely with NcoI or double digested with NcoI and EcoRV (NEB) overnight at 37 °C. RNAase (Sigma) was added to the reaction to avoid RNA contamination for later detection. Endonuclease digested gDNA was separated on a 0.8% (w/v) agarose gel for 4 h at 40 V in TAE buffer and imaged as described in Section 3.4.4. The gel was then washed with 0.25 M HCl with mild agitation (30 rpm) for 10 min to depurinate the DNA following by 15 min denaturation with 0.5 M NaOH and 20 min neutralisation with buffer 1 M Tris-HCl, 1.5 M NaCl, pH 7.5. After these washing steps, the DNA samples were transferred to a positively charged nylon membrane (Roche) through reverse capillary action overnight with 10 × saline-sodium citrate buffer (SSC) buffer. Transferred DNA fragments were covalently cross-linked to the membrane by UV cross-linking in a CL-100 (UVP) UV crosslinker at 1200 mJoules. The membrane was

pre-incubated with 20 mL DIG Easy Hyb™ Granules solution (Roche) at 42 °C for 1 h in a hybridisation oven (Techne™ Hybridisation Oven). The DIG-labelled probe (Section 3.4.17.1) was denatured for 5 min at 100 °C following by a rapid cool down on ice prior to mixing with 20 mL DIG Easy Hyb™ Granules solution. The membrane was incubated overnight with this hybridisation solution with 20 µL denatured probe. Following the hybridisation, the membrane was washed twice with low-stringency condition (42 °C, 5 min, 1 × SSC with 0.01% (w/v) SDS) and high-stringency condition (65 °C, 15 min, 0.5 × SSC with 0.01% (w/v) SDS) sequentially. The blot was then developed with DIG wash and block buffer set (Roche) following the manufacturer's instructions. In general, the blot was equilibrated in 1 × wash buffer for 5 min at room temperature before being blocked in blocking buffer for 30 min. The anti-DIG AP-conjugate antibody (Roche) was then added to the blocking buffer with a 1:10,000 dilution and the blot was incubated further for 30 min followed by 15 min washing with 1 × wash buffer for twice. The membrane was placed in a plastic folder and chemiluminescent substrate CSPD (Roche) was applied and incubated for the detection. The blots were exposed onto Amersham Hyperfilm™ ECL film for 1-30 min and developed with a KODAK film developer. For stripping the blot, 0.4 M NaOH was applied to the membrane and washed for 5 min at 42 °C in the hybridisation oven for twice. The blot was then washed with 1 x SSC buffer for 10 min at 42 °C for three times before re-probing.

Table 3.15 Amplification DIG-labelled probes.

Reaction mix			
Reagent	Volume (µl) DIG-labelled probe	unlabelled control probe	Final quantity
10× PCR buffer with $MgCl_2$	5	5	1×
PCR DIG probe synthesis mix	5	-	200 µM dATP, dCTP, dGTP, 130 µM dTTP, 70 µM DIG-dUTP
dNTP stock solution	-	5	200 µM each dNTP
10 µM forward primer	5	5	1 µM
10 µM reverse primer	5	5	1 µM
Enzyme mix	0.75 µl	0.75 µl	
DNA template (100 pg/µL)	1 µl	1 µl	100 pg

Thermal cycles			
Step	Temperature (°C)	Time (sec)	Number of cycles
Initial denaturation	95	120	1
Denaturation	95	30	20

Annealing	60	30	
Elongation	72	40	
Final elongation	72	420	1

3.4.18 CRISPR-Cas9 mediated gene editing

3.4.18.1 Design of sgRNA

Single-guide RNA (sgRNA) sequences used for CRISPR-Cas9 gene editing were designed using following criteria: i, sgRNA sequence were chosen besides the protospacer adjacent motif (PAM, 5′-NGG-3′) with a general high melting temperature (T_m); ii, the GC content of the 6 nucleotides preceding the PAM site were at least 50%; iii, there should not be any TT or GCC motifs ahead of the PAM site. The sgRNA sequence was then BLAST searched against *T. brucei* genome to minimise any off-target effect.

3.4.18.2 PCR amplification of sgRNA

To generate double stranded sgRNA, PCR reactions were set up as described in (Table 3.16) with OL39 and OL40 primers using Q5® High-Fidelity DNA polymerase (NEB). OL39 was a gene specific primer while OL40 was a universal primer used for *T. brucei* CRISPR Cas9 editing (Rico *et al.*, 2018). The success of the annealing of the primers was checked on a 1% (w/v) agarose gel.

Table 3.16 Amplification of sgRNA for CRISPR Cas9.

Reaction mix			
Reagent		Volume (µl)	Final quantity
5× Q5 reaction buffer		8	1×
10 µM gene-specific forward primer		8	2 µM
10 µM G00 reverse primer (sgRNA scaffold)		8	2 µM
10 mM dNTP mix		0.8	0.2 mM
Q5 HiFi Polymerase		1.2	2.4 U
ddH_2O		to 40 µl	
Thermal Cycles			
Step	Temperature (°C)	Time (sec)	Number of cycles
Initial denaturation	98	30	1
Denaturation	98	10	5

Annealing	50	30
Elongation	72	15

3.5 Protein biochemistry

3.5.1 Stable isotope labelling in cell culture (SILAC) of *T. brucei* BSF and PCF

For stable isotope labelling in cell culture of *T. brucei* BSF cells, HMI-11T medium used for culturing was prepared as described in (Table 3.1) with changes in the following components: i, the usage of dialysed heat-inactivated FBS (PAA labs); ii, the addition of folic acid (Sigma); iii, the addition of stable isotope labelled L-arginine:HCl (^{13}C6) (R6) and L-lysine:HCl 4(^{13}C4) (K4). For PCF, SDM-79 medium was prepared as described in (Table 3.1) with the changing to dialysed heat-inactivated FBS and addition of stable labelled R6K4 amino acids.

The SILAC labelling of the cells were performed as described (Urbaniak *et al*., 2012). In general, for *T. brucei* BSF, the cells were kept in culture in the HIM11-T + R0K0 medium for several divisions to let the cell adapt to the relatively low Arg and Lys medium. The BSF WT cells were then kept in culture with the HMI11-T + R0K0 medium while the mutant cell lines were cultured with HMI11-T + R6K4 medium. Cells were cultured for nine divisions to allow full labelling to steady-state and to reach mid log phase. For *T. brucei* PCF, cells were inoculated directly to the SDM-79 + R0K0 or + R6K4 medium and nine divisions time were allowed to achieve full labelling to steady-state.

3.5.2 *T. brucei* cell lysate preparation

For general cell lysate preparation of *T. brucei*, $5 \times 10^6 – 1\times 10^7$ log phase of *T. brucei* BSF and PCF cells were harvested and washed twice with 1 × TDB buffer (BSF) or 1 × PBS (PCF) cells at 4°C. After the last wash, the cells were transferred to 1.5 mL Eppendorf tube and pelleted for 10 s at 16,000 g following by re-suspension in equal volume of corresponding buffer and 2 × sodium dodecyl sulphate polyacrylamide (SDS) sample buffer (Table 3.16) with addition of 100 mM DTT at final concentration. The cell lysate was then heated up at 50°C for 20 min before further analysis. After cooling down, nuclease (Pierce™ Universal Nuclease for Cell Lysis) was added to samples to decrease the DNA viscosity.

Table 3.17 4 × SDS sample buffer for Licor.

Composition	pH	Amount
Tris-HCl		0.666 g
EDTA		0.006 g
SDS	8.5	0.4 g
Glycerol		4 g
Orange G		0.02 g
ddH_2O		to 10 mL

3.5.3 Cross-linking using DTSSP

For cross-linking reactions, 2×10^8 or 1×10^9 of *T. brucei* PCF and BSF cells were harvested at 600 g or 800 g, respectively, 4 °C for 10 min and subjected to hypotonic lysis with water containing 0.1 mM N_α-Tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK), 1 µg /µL leupeptin, 0.1 mM phenylmethylsulphonylfluoride (PMSF) at 1×10^9 cells/mL. After incubation on ice for 5 min, the supernatants containing soluble cytosolic proteins were removed by centrifugation at 16,000 g for 5min, 4C, except for the TbGPI3 cross-linking reaction where some experiments were carried out using the whole hypotonic cell lysate. The pelleted membrane fractions (or the whole cell lysates) were then re-suspended in 196 µL cross-linking reaction buffer (20 mM NaHEPES pH 7.8, 25mM KCl, 2 µg/µL E-64 (Sigma), 1 µg/µL leupeptin) containing different concentrations of cross-linker DTSSP (3,3'-dithiobis(sulfosuccinimidyl propionate)) (Thermo). The reactions were carried out for 1 h on ice and quenched by adding Tris-HCl buffer (pH8.0) to a 20 mM final concentration. The membranes were then pelleted (16,000 g, 10 min, 4C), resuspended in 1 mL washing buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl) pelleted again and solubilised in 196 µL lysis buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl, 2% TX-100) containing protease inhibitors (0.1 mM TLCK, 1 µg /µL leupeptin, 1 µg /µL aprotinin, 0.1 mM PMSF) and immunoprecipitation was performed described in Section 3.5.4.

3.5.4 Immunoprecipitation

For immunoprecipitation, cells were harvested at 600 g (for PCF) or 800 g (for BSF), 4 °C for 10 min and then lysed in 1 mLlysis buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl containing different detergents; 0.5% digitonin, 1% digitonin, 1% Triton X-100

(TX-100), 1% n-octyl-beta-glucoside (NOG) or 1% decyl-β-D-maltopyranoside (DM)) with final concentration equivalent to 1×10^9 cells/mL. All buffers contained 1 mM TLCK, 1 µg /mL leupeptin, 1 µg /mL aprotinin, 1 mM PMSF as protease inhibitors. After centrifugation at 16,000 g, 4 °C for 20 min, supernatants were collected and incubated with MycTrap™ agarose beads (Chromotek) or anti-HA magnetic beads (Thermo) for 1 h at 4 °C. The beads were washed three times in the 1 mL washing buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl containing the corresponding detergents) and the bound proteins were eluted under either denaturing or native conditions. The denaturing elution condition used in this study was heating the beads in 30 µL1 × SDS sample buffer at 50 °C for 20 min in the presence of reducing agent (100 mM dithiothreitol (DTT)) and the native elution condition was to elute the beads three times with 10 µL 0.5 mg/mL c-Myc peptide (Thermo) or 1 mg/mL HA peptide (Sigma) in the corresponding detergent containing buffer. The denatured eluates or the combining native eluates for each sample were subjected to either SDS-PAGE (Invitrogen) or NativePAGE (Invitrogen) and further analysed either by Western blotting or proteomics.

3.5.5 SDS-PAGE and Native-PAGE

Protein samples from cell lysates and from the denaturing elution of tag-affinity beads beads were subjected to SDS-PAGE using various (10%, 12%, 4-12 %) pre-cast NuPAGE™ Bis-Tris gels (Invitrogen) gels with 1 × NuPAGE™ MOPS SDS Running Buffer (Invitrogen). Electrophoresis was performed in an XCell SureLock™ Mini-Cell tank (Invitrogen) at 200 V using a Power Pac 300 (BioRad). Aliquots (2 µL) of Precision Plus Protein™ All Blue Prestained Protein Standards (BioRad) were loaded alongside the protein samples to serve as an indication for the protein sizes on the gel. Protein samples eluted from beads under native conditions were separated on 3-12% gradient NativePAGE™ Bis-Tris gels (Invitrogen) following the manufacture's instruction with electrophoresis with 1× dark blue NativePAGE Cathode Buffer (Invitrogen) at 150 V for 30 min and then changed to 1 × light blue NativePAGE Cathode Buffer running at 150 V for another 60-80 min until the desired separation for proteins was achieved. In the process of electrophoresis, cathode buffers were filled in the inner chamber of the XCell SureLock™ Mini-Cell tank while the 1 × NativePAGE running buffer was filled in the outer chamber.

3.5.6 Western blotting

3.5.6.1 **Generation of anti-Myc magnetic beads**

For generation of home-made anti-Myc magnetic beads, 1 mL NHS-activated magnetic beads (Pierce™) were washed in ice cold wash buffer (1 mM hydrochloric acid) once and re-suspended with 262 µL coupling buffer (50 mM sodium tetraborate decahydrate (pH 8.5)). Equal volume of anti-Myc mouse monoclonal antibody (0.38 mg/mL) gained from University of Dundee MRC PPU reagent unit was added to the magnetic beads and incubated at 4 °C overnight. After the reaction, the supernatant was removed with the help of magnetic stand and 1 mL quenching buffer (3 M ethanolamine pH 9.0) was added to the beads to quench the reaction. The anti-Myc magnetic beads were washed three times with 0.1 M glycine buffer (pH 2.0) and stored in 1 × PBS buffer for future usage.

After separation by SDS-PAGE or NativePAGE, proteins were transferred onto either nitrocellulose membrane or PVDF using a iBlot™ 2 Gel Transfer Device (Invitrogen) at 25 V for 7-10 min according to the manufacturer's instructions. The membrane was then incubated with Ponceau S buffer (Sigma) for 5 min following de-staining in water for 5 min, three times. to verify the success of protein transfer. This also served as a quantitative control for protein loading. Blocking of the membrane was achieved using SNAP i.d. ® 2.0 Protein Detection System for Western Blotting (Merck) according to manufacturer's instruction with the application of 30 mL blocking buffer (50mM Tris-HCl pH7.4, 0.15M NaCl, 0.25% BSA, 0.05% (w/v) Tween-20, 0.05% NaN_3 and 2% (w/v) Fish Skin Gelatin). This step was to prevent non-specific binding of antibodies introduced in the following steps. After blocking, the membrane was incubated with primary antibody diluted in blocking buffer to an appropriate concentration for 1 h at room temperature with a gentle rotation. Relevant dilutions of primary antibodies are indicated specifically for each experiment in the results chapters and/or in (Table 3.18). The membrane was then transferred to the iBlot assembling cassette and washed three times with 30 mL 1 × phosphate buffered saline plus Tween (PBST) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, 0.1 % Tween 20) buffer before incubation with IRDye® 680RD and 800CW conjugated secondary antibodies or horse radish peroxidase (HRP)-conjugated secondary antibodies. The primary and secondary antibodies used for Western blotting in this study were listed in (Table 3.18). The membranes were imaged either by LI-COR Odyssey infrared imaging system (LICOR Biosciences, Lincoln, NE) or incubated with the ECL Plus Western Blotting detection

reagent (GE Healthcare) according to the manufacturer's instruction and exposed to ECL Hyperfilm (GE Healthcare) and developed with a KODAK film developer.

Table 3.18 List of antibodies used for Western blotting.

Antibodies	Dilution	Suppliers
Anti-Myc rat monoclonal antibody	1:1,000	Chromotek, 9E1
Anti-HA rat monoclonal antibody	1: 500/1:1,000	Sigma, 3F10
IRDy 800CW conjugated goat anti-rat IgG antibody	1:15,000	LICOR Biosciences
IRDy 680RD conjugated goat anti-mouse IgG antibody	1:20,000	LICOR Biosciences
HRP-conjugated anti-rat antibody	1:3,000	Sigma, A9037

3.5.7 PNGase F digestion

To set up the Peptide-N-Glycosidase F (PNGase F) digestion, 5×10^7 BSF WT and Tb927.11.11740-3HA expressing cells were harvested and immunoprecipitation was performed as described in Section 3.5.4 until the end of the washing steps. The on-bead PNGase F digestion was then set up using reagents from the PNGase F kit (NEB). The beads from the IPs with BSF WT or Tb927.11.11740-3HA lysates were denatured in 10 µL 1 × Glycoprotein Denaturing Buffer in the absence of 10 mM DTT at 50 °C for 10 min. The PNGase F digestion was then performed on the denatured samples in the reaction buffer (1× GlycoBuffer 2) containing 1 % NP-40 with or without 1 µL PNGase F (NEB).

3.6 Immunofluorescence microscopy (IFM)

3.6.1 Cell fixing, staining, and imaging

All the solutions used in this protocol were filtered either on a 0.2 µm Minisart filtering unit (Sartorious) or with TPP® Filtermax rapid bottle filter unit (Merck). Aliquots of 1×10^7 of *T. brucei* BSF or PCF cells were harvested at 800 g or 600 g, 4 °C for 10 min and washed once with ice-cold 1 × TDB or 1 × PBS buffer, respectively. The cells were then resuspended in 0.5 mL of TDB or PBS and transferred to 1.5 mL Eppendorf tubes containing 0.5 mL 8 % paraformaldehyde (PFA) in either TDB or PBS to achieve a final concentration of 5×10^6 cells/mL and 4% paraformaldehyde. The cells were fixed for 30 min on ice and 10 µL aliquot of the cells were dried on 12-well diagnostic microscope slides (Thermo). The slides were pre-treated with Poly-L-Lysine solution

(Sigma) prior to the cell coating to enhance cell attachment. The cells fixed on the 12-well slides were rehydrated in 1 × PBS buffer for 5 min and then permeabilised with 1 × PBS containing 0.5% TX-100 for 15 min. The slides were washed three times with 1 × PBS followed by applying the blocking buffer for 1 h at room temperature, the composition of the blocking buffer is given in (Table 3.18). After the blocking step, the slides were washed with 1 × PBS buffer for 5 min, twice, and then incubated with primary antibodies diluted in blocking buffer for 1 h at room temperature. The slides were washed three times with 1 × PBS buffer and incubated with secondary antibodies diluted in the blocking buffer for 1 h at room temperature in the dark. The antibodies and corresponding dilutions for IFM used in this study were listed in Table (3.20). For both primary antibodies and secondary antibodies, if more than one antibody was used, they were diluted in the same dilution, buffer and centrifugated at 16000 × g, 4°C for 10 min before applying to the slides. ProLong® Gold antifade reagent with DAPI (Invitrogen) was spotted on the slide and cover glasses (VWR) were lowered slowly on the drop of the mounting agent. The slides were left curing overnight in the dark at room temperature before being sealed with nail varnish. Cells were analysed by a Zeiss LSM 700 confocal microscope and all the images were processed using Fiji v. 2.0.0 (Schindelin *et al.*, 2012).

Table 3.19 Blocking buffer for IFM.

Composition	Amount	Final quantity	Supplier
Fish skin gelatin	0.5 g	5% (w/v)	Sigma
Normal goat serum	1 mL	10 %	Sigma
Sodium Azide	25 µL	0.05% (v/v)	Sigma
10% TX-100	50 µL	0.05% (w/v)	Sigma
1 × PBS	9 mL	90 % (v/v)	Sigma

Table 3.20 List of antibodies used for IFM.

Antibodies	Dilution	Suppliers
Anti-Myc mouse monoclonal antibody	1:1,000	Cell signalling
Anti-HA rat monoclonal antibody	1: 500	Sigma, 3F10
Rabbit anti-BiP	1:2,000	Kind gift from Jay Bangs (University of Buffalo)
Alexa Fluor® 488- conjugated goat anti-mouse	1:1,000	Invitrogen
Alexa Fluor® 488-conjugated goat anti-rat	1:1,000	Invitrogen
Alexa Fluor® 594-conjugated goat anti-rabbit	1:1,000	Invitrogen

3.7 Proteomics

3.7.1 Label free quantitative proteomics

3.7.1.1 Sample preparation and data acquisition

BSF WT and mutant cell lines, PCF WT and mutant cell lines were cultured and 2×10^8 – 1×10^9 cells of each cell line were harvested, washed in washing buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl) and lysed in 1 mL lysis buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl containing different detergents). The detergent conditions are indicated specifically for each experiment in the results chapter. After centrifugation at 16,000 g, 4 °C for 20 min. The supernatants were mixed with either 20 µL Myc-Trap™ (Chromotek) or 20 µL anti-HA magnetic beads (Thermo) and incubated for 1 h at 4 °C. The beads were washed three times in the 1 mL same buffer and bound proteins were eluted with 1 × SDS sample buffer by heating at 50 °C for 20 min and subjected to SDS-PAGE, running the proteins only 10 cm into the gel. Whole lanes containing WT and mutant cell lines samples were cut identically into 3-4 slices and the gel pieces were dried in Speed-vac (Thermo Scientific) for in-gel reduction with 0.01M dithiothreitol (DTT) and alkylation with 0.05M iodoacetamide (Sigma) for 30 min in the dark. The gel slices were washed in 0.1M NH_4HCO_3, and digested with 12.5 µg/mL modified sequence grade trypsin (Roche) in 0.02M NH_4HCO_3 for 16 h at 30˚C. Samples were dried and re-suspended in 50 µL 1% formic acid and the liquid phase subjected to liquid chromatography on Ultimate 3000 RSLC nano-system (Thermo Scientific) fitted with a 3 Acclaim PepMap 100 (C18, 100 µM× 2 cm) trap column and then separated on an Easy-Spray PepMap RSLC C18 column (75 µM× 50 cm) (Thermo Scientific). Samples (15 µL) were loaded in 0.1% formic acid (buffer A) and separated using a binary gradient consisting of buffer A and buffer B (80% acetonitrile, 0.1% formic acid). Peptides were eluted with a linear gradient from 2 to 35% buffer B over 70 min. The HPLC system was coupled to a Q-Exactive Plus Mass Spectrometer (Thermo Scientific) equipped with an Easy-Spray source with temperature set at 50˚C and a source voltage of 2.0 kV. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier Project PXD022979.

3.7.1.2 **Protein identification by MaxQuant**

RAW data files were analysed using MaxQuant version 1.6.10.43, with the in-built Andromeda search engine (Tyanova, Temu and Cox, 2016), using the *T. brucei brucei* 927 annotated protein sequences from TriTrypDB release 46 (Aslett *et al.*, 2010) supplemented with the *T. brucei brucei* 427 VSG221 (Tb427.BES40.22) protein sequence. The mass tolerance was set to 4.5 ppm for precursor ions and MS/MS mass tolerance was set at 20 ppm (MaxQuant default parameters). The enzyme was set to trypsin, allowing up to 2 missed cleavages. Carbamidomethyl on cysteine was set as a fixed modification. Acetylation of protein N-termini, and oxidation of methionine were set as variable modifications. Match between runs was enabled, allowing transfer of peptide identifications of sequenced peptides from one LC-MS run to non-sequenced ions with the same mass and retention time in another run. A 20-min time window was set for alignment of separate LC-MS runs. The false-discovery rate for protein and peptide level identifications was set at 1%, using a target-decoy based strategy.

3.7.1.3 **Data analysis**

Data analysis was performed using custom Python scripts, using the SciPy ecosystem of opensource software libraries (Virtanen *et al.*, 2020). The data analysis pipeline is available at GitHub https://github.com/mtinti/TbGPI3 and Zenodo https://zenodo.org/record/4310034 repositories, DOI:10.5281/ zenodo.3735036. The MaxQuant proteinGroups.txt output file was used to extract the iBAQ scores for forward trypanosome protein sequences identified with at least two unique peptides and with an Andromeda score >4. The protein iBAQ scores were normalised for sample loading by dividing each iBAQ value by the median of all the iBAQ values in each experiment. Missing values were replaced by the smallest iBAQ value in each sample. Differential abundance analysis between the bait and control samples was performed with the ProtRank Python package (Medo, Aebersold and Medová, 2019). Briefly, ProtRank performs a rank test between each control and bait sample pair to output as signed-rank and false discovery rate values. The signed-rank is proportional to the significance of the differential abundance of the protein groups between the bait and control samples. The BSF protein intensity (abundance) rank (Ji, Tinti and Ferguson, 2021) was computed from a recent dataset (Tinti *et al.*, 2019) published by our laboratory of *T. brucei* protein half-lives computed from a label-chase experiment. In

those experiments, BSF parasites were labelled to steady-state in medium SILAC culture medium (M) and then placed into light SILAC culture medium (L). Seven time points, with three biological replicates, were sampled and each mixed 1:1 with BSF lysate labelled to steady state in heavy SILAC culture medium (H) to provide an internal standard for normalisation. Each sample was then separated into 10 sub-fractions for LC-MS/MS (thus, a total of 210 LC-MS/MS analyses were performed). Here, we exploited the heavy-labelled internal standard in every sample: The log10 summed eXtracted Ion Currents (XICs) of the heavy-labelled peptides for each protein were averaged across the BSF replicates and used to rank a very deep BSF proteome from the most abundant (rank = 1) to the least abundant (rank = 7149). Undetected proteins were given a rank of 7150.

3.7.2 Cross-linking label free quantitative proteomics and data analysis

For label free quantitative proteomics identification, the cross-linking reaction for the whole cell lysates or for the cell membrane fractions were set up as described in Section 3.5.3 followed by immunoprecipitation indicated in Section 3.5.4. The sample processing was the same as the label free quantitative proteomics as introduced in Section 3.7.1.1. For protein identification by MaxQuant, a variable modification (145.02 Da) on lysine residue introduced by DTSSP cross linking was added when performing the MaxQuant search. The principle of reversible DTSSP cross-linking is shown in Fig 3.1.

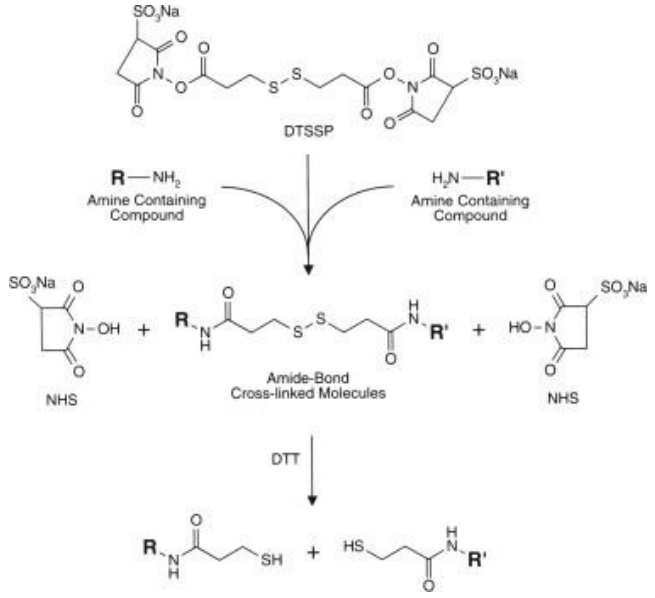

Figure 3.1 Peptide sequences (R and R_1) containing lysine residues, with side chain primary amino groups, react with DTSSP replacing the N-hydroxy succinimide (NHS)

leaving groups. The cross-linked peptides provide the covalent cross-linking between the proteins containing the R and R_1 peptides, thus allowing co-immunoprecipitation under stringent denaturing conditions. Once isolated, the cross-linked proteins can be separated from each other by reduction of the disulphide bond within the cross-liner with dithiothreitol (DTT). Image is taken from (Hermanson, 2013).

3.7.3 SILAC proteomics

3.7.3.1 **Sample preparation**

BSF WT and mutant cell lines were cultured in HMI11-T + R0K0 and HMI11-T + R6K4 medium (as described in section 3.5.1), respectively, while PCF WT and mutant cell lines were cultured in SDM-79 + R0K0 and SDM-79+ R6K4 medium respectively (as described in section 3.5.1). For each SILAC experiment, WT and mutant cells were count three individual times with haemocytometer and identical numbers of cells ($2 \times 10^8 – 1 \times 10^9$ cells, depending on the experiment) from the WT and mutant samples were mixed together. The cells were harvested and lysed in 1mL lysis buffer (50 mM Tris-HCl, pH7.4, 150 mM NaCl containing different detergents). The detergent conditions are indicated specifically for each experiment in the results chapter. After centrifugation at 16,000 g, 4 °C for 20 min. The supernatants were mixed with 20 µL Myc-Trap™ (Chromotek) and incubated for 1 h at 4 °C. The beads were washed three times in the same buffer and bound proteins were eluted with 1 × SDS sample buffer containing 100 mM DTT and boiled at 50 °C for 20 min and subjected to SDS-PAGE, running the proteins only one third of the gel. Whole lanes containing WT and mutant cell lines samples were cut identically into 3-4 slices and prepared for LC-MS/MS identification as described in (Section 3.7.1.1). Samples were dried and re-suspended in 50 µL 1% formic acid and then subjected to liquid chromatography on Ultimate 3000 RSLC nano-system (Thermo Scientific) fitted with a 3 Acclaim PepMap 100 (C18, 100 µM× 2 cm) trap column and then separated on an Easy-Spray PepMap RSLC C18 column (75 µM× 50 cm) (Thermo Scientific). Samples (15µL) were loaded in 0.1% formic acid (buffer A) and separated using a binary gradient consisting of buffer A and buffer B (80% acetonitrile, 0.1% formic acid). Peptides were eluted with a linear gradient from 2 to 35% buffer B over 124 min. The HPLC system was coupled to a OrbiTrap Velos Pro (Thermo Scientific) equipped with an Easy-Spray source with temperature set at 50˚C and a source voltage of 7.9 kV.

3.7.3.2 **Data processing**

For the SILAC experiments the RAW data files were analysed using MaxQuant version 1.5.8.3, with the in-built Andromeda search engine (Tyanova, Temu and Cox, 2016), using the T. brucei brucei 927 annotated protein sequences from TriTrypDB release 32 (Aslett *et al.*, 2010). The mass tolerance was set to 4.5 ppm for precursor ions and MS/MS mass tolerance was set at 20 ppm (MaxQuant default parameters). The enzyme was set to trypsin, allowing up to 2 missed cleavages. Carbamidomethyl on cysteine was set as a fixed modification. The heavy (R6K4,) and light (R0K0) modifications were used for SLAC quantification. Acetylation of protein N-termini, and oxidation of methionine were set as variable modifications. Match between runs was enabled, allowing transfer of peptide identifications of sequenced peptides from one LC-MS run to non-sequenced ions with the same mass and retention time in another run. A 20-min time window was set for alignment of separate LC-MS runs. The false-discovery rate for protein and peptide level identifications was set at 1%, using a target-decoy based strategy. The heavy over light SILAC log2 ratios were analysed with a t-test implemented in the scipy python package (Virtanen *et al.*, 2020) against the null hypothesis that the log2 ratio is zero. The minus log10 of the p-values were plotted against the mean log2 ratios. For the experiments without biological replicates the log ratio was visualized against the log10 Score of the MaxQuant output.

3.8 *T. brucei* cell free system and radiolabelling

3.8.1 *T. brucei* cell free system preparation

The *T. brucei* BSF cell free system for different cell lines used in this study was prepared as described previously (Güther, Masterson and Ferguson, 1994). Briefly, cells were lysed in buffer A (0.1 mM TLCK, 1 µg/mL leupeptin) for 5 min on ice, an equal volume of buffer B (0.1 M NaHEPES, 50 mM KCl, 10 mM $MgCl_2$, 20% glycerol, pH 7.4) were added, and the cell lysates could be kept in - 80°C for later use. Aliquots of trypanosome lysate (40 µL, equivalent to 4×10^7 cells) were incubated with 0.5 µCi GDP-[^{3}H]Man (American Radiochemical Company, 20-30 Ci/mmol) and 2 mM UDP-GlcNAc for 20 min at 37 °C in the incorporation buffer (50 mM HEPES, 25 mM KC1, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 0.5 mM dithiothreitol, 1.25 µg/ml tunicamycin,). In chase experiments, 1 mM cold GDP-mannose was added and incubated for 10 min at 37 °C. The reactions were terminated by the addition of 0.2 mL of chloroform, methanol (1:1,

v/v) and lipids were extracted in the solvent overnight. The lipid extracts were dried under a stream of nitrogen, resuspended in 50 µL of water saturated with 1-butanol, and extracted three times with 50 µL of 1-butanol saturated with water. The combined upper phases were washed twice with 100 µL of water saturated with 1-butanol. The upper 1-butanol phase were dried under N_2 and re-suspend in 20 µL 1-butanol. Aliquots of the washed lipid extract were analysed by HPTLC.

3.8.2 High Performance Thin layer chromatography (HPTLC)

Radiolabelled samples were applied on Silica Gel 60 aluminium backed HPTLC plates (Merck) and developed with solvent (chloroform, methanol, 1 M ammonium acetate, 13 M ammonia, water (180:140:9:9:23, v/v)). The HPTLC plates were fluorographed by spraying with En3hance® Spray solution (Perkin Elmer) and incubated for overnight or several weeks at -80°C to expose with film (Amersham Hyperfilm MP, GE Healthcare) in the presence of an intensifying screen (Kodak). Films were developed with a KODAK film developer. For purification of radiolabelled GPI glycolipids A' and θ used in Section 6.9, the Rf values of these two products were first determined in a small scale experiment followed by a bulk radiolabelling of BSF WT cell free system as described in Section 3.8.1, except that 120 µl of cell free system and 1.5 µCi of GDP-[^{3}H]Man were used. After separation the radiolabelled GPI glycolipids mix on an HPTLC plate, the regions that contained glycolipids A' and θ, which were estimated based on their Rf values, were scraped and the glycolipids were recovered by butanol extractions as described in Section 3.8.1.

4 Results I – Studies on *Trypanosoma brucei* GlcNAc-Phosphatidylinositol De-N-acetylase (TbGPI12)

4.1 Identification by SILAC proteomics of proteins that co-immunoprecipitated with TbGPI12

4.1.1 Analysis of TbGPI12 Gene Product.

The gene *TbGPI12* (Tb927.11.12080) has been shown to encode the de-N-acetylase that catalyses the second step in the *T. brucei* glycosylphosphatidylinositol biosynthesis pathway (GPI biosynthesis pathway), converting D-GlcNAcα1–6-D-*myo*-inositol-1-HPO_4-*sn*-1,2-diacylglycerol (GlcNAc-PI) to D-GlcNα1–6-D-*myo*-inositol-1-HPO_4-*sn*-1,2-diacylglycerol (Chang *et al.*, 2002). A conditional null mutants for this gene have shown that it is essential for the survival of bloodstream form but not procyclic form *T. brucei* (Chang *et al.*, 2002; Güther *et al.*, 2006). TbGPI12 is predicted to be 252 amino acids in length, with a theoretical molecular mass of 28.6 kDa, and to contain a transmembrane domain at its C-terminus using the Phobius web server (Käll, Krogh and Sonnhammer, 2007). The enzymes of GPI anchor biosynthesis pathway have been shown to be generally of low abundance, with few of them appearing in the proteomic studies performed on BSF and PCF parasites (i.e., several do not appear TriTrypDB proteomics summary data) (Aslett *et al.*, 2010) and see Table 1.1. Consistent with this theme of general low-abundance, and according to recent proteomics work from our lab that orders proteins based on their relative abundance (Tinti, Maria Lucia S Güther, *et al.*, 2019), TbGPI12 ranks 5818th out of 7148 detected protein groups in bloodstream form *T. brucei*.

Previous studies (Smith *et al.*, 1996) have shown that the addition of synthetic GlcNAc-PI to a BSF cell-free system in the presence of GDP-[^{3}H]Man primes the formation of radiolabelled downstream GPI intermediates far more efficiently than the addition of synthetic GlcN-PI. This suggests that there may be substrate channelling between TbGPI12 and the enzyme catalysing following step, α1-4 mannosyltransferase I, (TbGPI14, gene ID: Tb927.6.3300) and that TbGPI12 and TbGPI14 may form and work as a complex in the ER (Smith *et al.*, 1996). Therefore, we decided to look for this hypothetical complex and, while doing so, to optimise the conditions for

immunoprecipitation of low abundance ER transmembrane proteins in the GPI pathway. The first steps of GPI anchor biosynthesis pathway are identical in BSF and PCF parasites, so the co-immunoprecipitation of TbGPI12 could be performed in both forms of the parasite. Since PCF parasites are easier to culture to high density, we elected to use that life-cycle stage for these early studies.

4.1.2 Generation of anti-Myc magnetic beads and immunoprecipitation (IP) of PCF TbGPI12-Myc.

A *T. brucei* PCF cell line conditionally expressing TbGPI12-Myc previously generated in the lab (Güther *et al.*, 2006) was used in the following IP experiments. In order to get an efficient IP, home-made anti-Myc magnetic beads were generated by covalently linking a mouse anti-Myc monoclonal antibody to NHS-activated magnetic beads (Section 3.5.6.1). The purity of the anti-Myc antibody was checked before coupling to the beads by SDS-PAGE and Coomassie blue staining, and the success of coupling the anti-Myc antibody to the magnetic beads was confirmed by identifying the antibody heavy and light chains eluted (by boiling for 20 min using SDS sample buffer) from the beads after coupling and washing (Fig 4.1A). The coupled anti-Myc magnetic beads were then used for immunoprecipitation of TbGPI12-Myc. In this experiment, 3 sequential IPs from 1% TritonX-100 (TX-100) cell lysates of induced (+Tet) and uninduced (-Tet) PCF *TbGPI12* conditional null mutant trypanosomes were performed, and SDS sample buffer eluted proteins were analysed by Western blot using rat anti-Myc antibody. TbGPI12-Myc was successfully detected in all 3 IPs from the +Tet samples, but not the –Tet samples (Fig 4.1 B).

It is clear from the TbGPI12-Myc data that the anti-Myc magnetic beads could IP the low abundance transmembrane protein TbGPI12-Myc, however the presence of TbGPI12-Myc in all three sequential IPs suggests that it is difficult to achieve quantitative IP of TbGPI12-Myc. This may be due to the use of a single Myc-tag and/or the very low concentration of the target TbGPI12-Myc protein in the whole protein lysates.

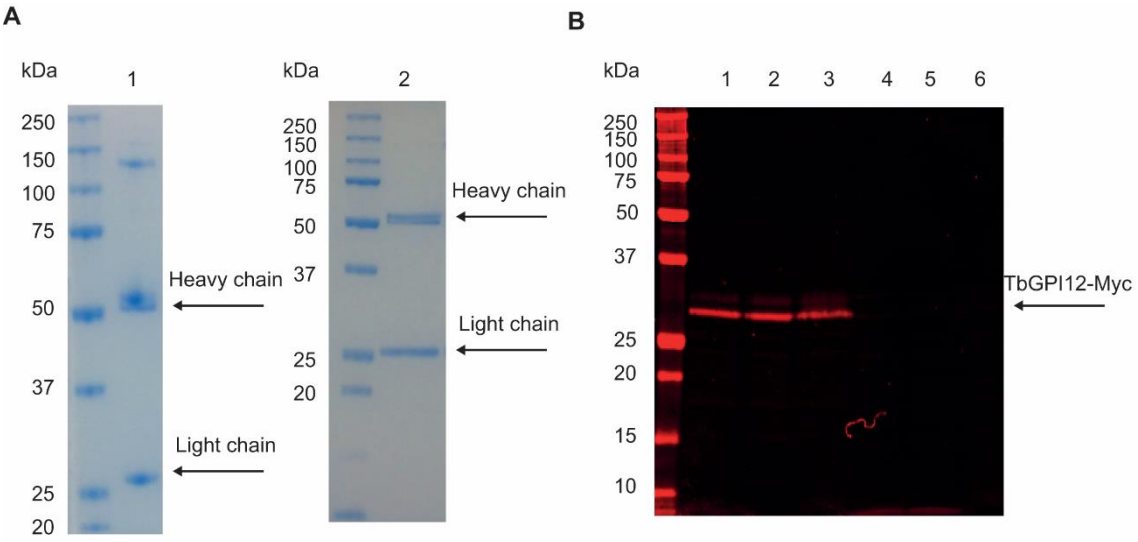

Figure 4.1 Coupling anti-Myc monoclonal antibody to NHS-activated magnetic beads and immunoprecipitation and Western blot of TbGPI12-Myc.

A: Coomassie blue stained SDS-PAGE gel of the anti-Myc monoclonal antibody before (lane 1) and after (lane 2) coupling to NHS-activated magnetic beads. After coupling, the antibody chains were eluted from the beads with SDS-sample buffer. B: Cell lysates (in buffer containing 1% TX-100) from the PCF TbGPI12 conditional null mutant Tet induced (lanes 1-3) and uninduced (lanes 4-6) were subjected to 3 sequential IPs, and the eluted proteins detected with rat anti-Myc antibody. The positions of molecular weight markers are shown on the left of each gel and blot.

4.1.3 Optimisation of detergent lysis and protein elution conditions for TbGPI12-Myc IPs.

To try to maintain any membrane complexes that TbGPI12-Myc may form, different concentrations of digitonin were explored for lysing the cells and compared to the use of 1% TX-100. Digitonin is recognised as a much milder detergent than TX-100 and is more suitable for solubilising intact membrane protein complexes. Western blot after IP of TbGPI12-Myc using different lysing conditions indicates that 0.3%, 0.5% and 1% digitonin all supported the extraction of TbGPI12-Myc, albeit with lower efficiency than 1% TX-100 (Fig 4.2 B). Out of these conditions, 0.5% digitonin was selected as the lysis condition for subsequent experiments. However, Ponceau staining of the same blot (Fig 4.2 A) showed that the use of SDS sample buffer for elution of the anti-Myc magnetic beads after IP was eluting many non-specifically bound proteins from the beads which would likely interfere with subsequent Nano-LC-MS/MS detection of the co-IP proteins with TbGPI12-Myc. Therefore, two different mild elution buffers: low

pH sodium citrate buffer and c-Myc peptide (Sigma) were tested. Low pH sodium citrate gave a less efficient elution than SDS sample buffer but a similar Ponceau staining background (Fig 4.2 C & D), while the c-Myc peptide buffer in this experiment did not elute any TbGPI12-Myc (Fig 4.2 D, lane 2). This result indicates that although home-made coupled anti-Myc magnetic beads could achieve an efficient IP for TbGPI12-Myc, it could also give a high background binding of many non-specific proteins, despite using mild detergent lysis and mild (low pH) elution conditions. While the background of non-specific proteins was low in the c-Myc peptide elution (Fig 4.2 C, lane 2), there was no specific elution of TbGPI12-Myc either. This was investigated further in section 4.1.4.

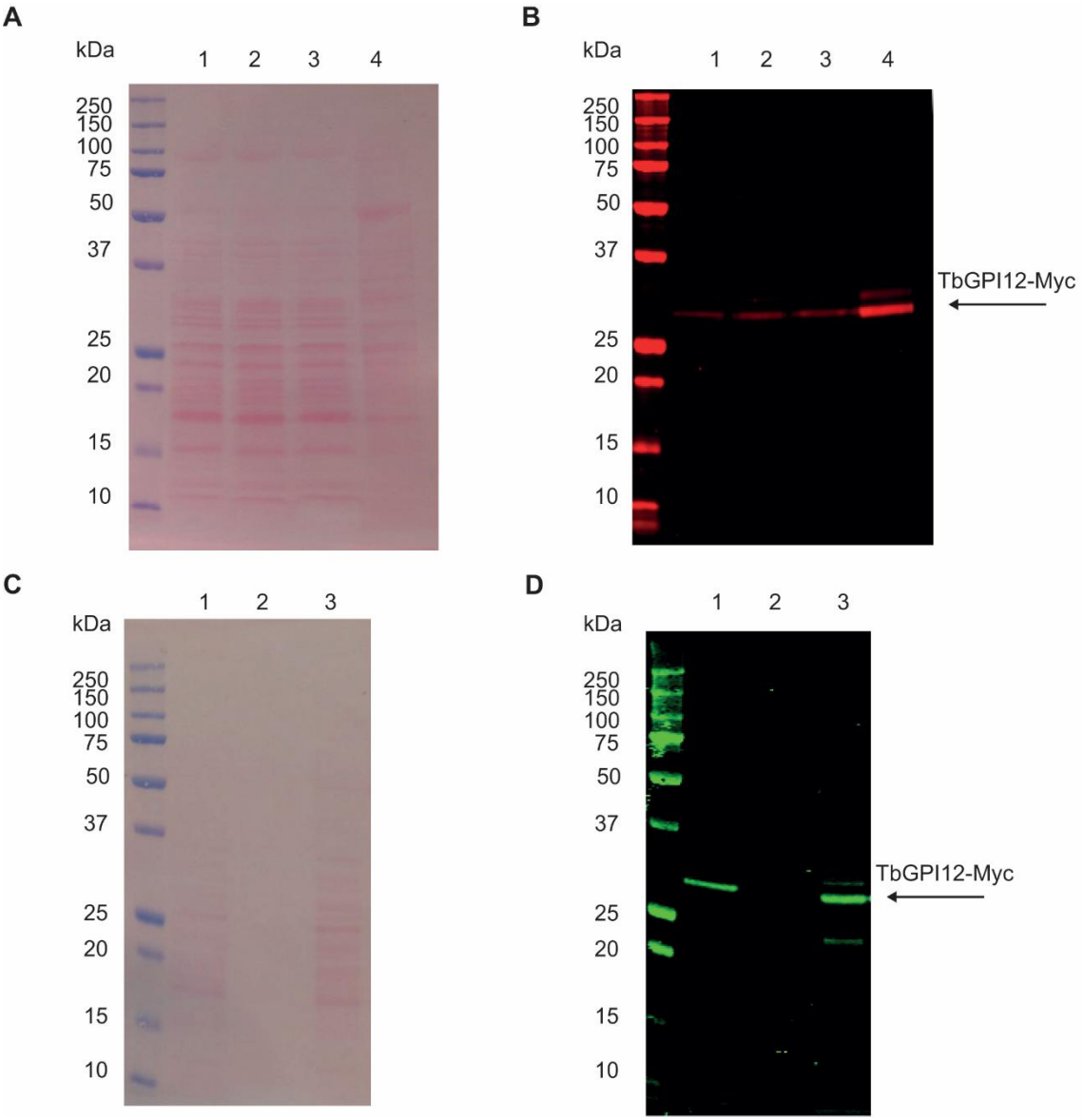

Figure 4.2 Test of detergent extraction conditions and elution conditions for IP of TbGPI12-Myc.

A and B: TbGPI12-Myc lysates prepared using different detergents were immunoprecipitated and by anti-Myc magnetic beads and eluted proteins were analysed by Ponceau staining and Western blot detection using rat anti-Myc antibody. Lane1-3: 0.3 %, 0.5 %, 1% digitonin, lane 4: 1% TX-100. C and D: Ponceau staining and anti-Myc Western blot of proteins eluted by different eluting buffers from anti-Myc magnetic beads. Lane 1: sodium citrate buffer pH 2.8, lane 2: 0.5 mg/ml c-Myc peptide (Sigma) in lysis buffer, lane 3: SDS sample buffer. All TbGPI12-Myc expressing cell lysates were extracted with 0.5% digitonin.

4.1.4 Standardisation of Myc-Trap™ beads for IP and selection of an alternative source of c-Myc peptide for specific elution.

The high Ponceau background staining observed using anti-Myc coupled magnetic beads raised concerns with respect to the higher possibility of detecting "false positive" co-IP proteins with TbGPI12-Myc in later proteomics experiments, even when including samples IPs from wild type cell line lysates as a negative control. Therefore, Myc-Trap™ agarose beads were tested as an alternative for for IP as Myc-Trap™ uses an anti-Myc nanobody engineered to reduce non-specific protein-protein binding, according to the manufacturer. To test the efficiencies of the Myc-Trap™ agarose beads, equivalent amounts of 0.5% digitonin cell lysates containing TbGPI12-Myc were immunoprecipitated using Myc-Trap™ beads in parallel with our home-made anti-Myc coupled magnetic beads. Proteins were then eluted from the beads using low pH sodium citrate buffer and analysed by Western blotting. As shown in Fig 4.3 A and B, the yield of TbGPI12-Myc recovered from Myc-Trap™ beads was higher, while the Ponceau staining background was lower. Meanwhile, a new c-Myc peptide (from Thermo instead of Sigma) was tested as an eluant using Myc-Trap™ beads for IP. The Western blot shown in Fig 4.3 D suggests that Thermo c-Myc peptide elution buffer is more efficient than low pH sodium citrate.

From the data described above, an IP procedure using 0.5% digitonin for lysing cells, Myc-Trap™ agarose beads for IP and c-Myc peptide from Thermo for eluting was selected for future TbGPI12-Myc IP experiments. In addition, we decided that these conditions could also be applied to IP other low abundant Myc-tagged transmembrane proteins.

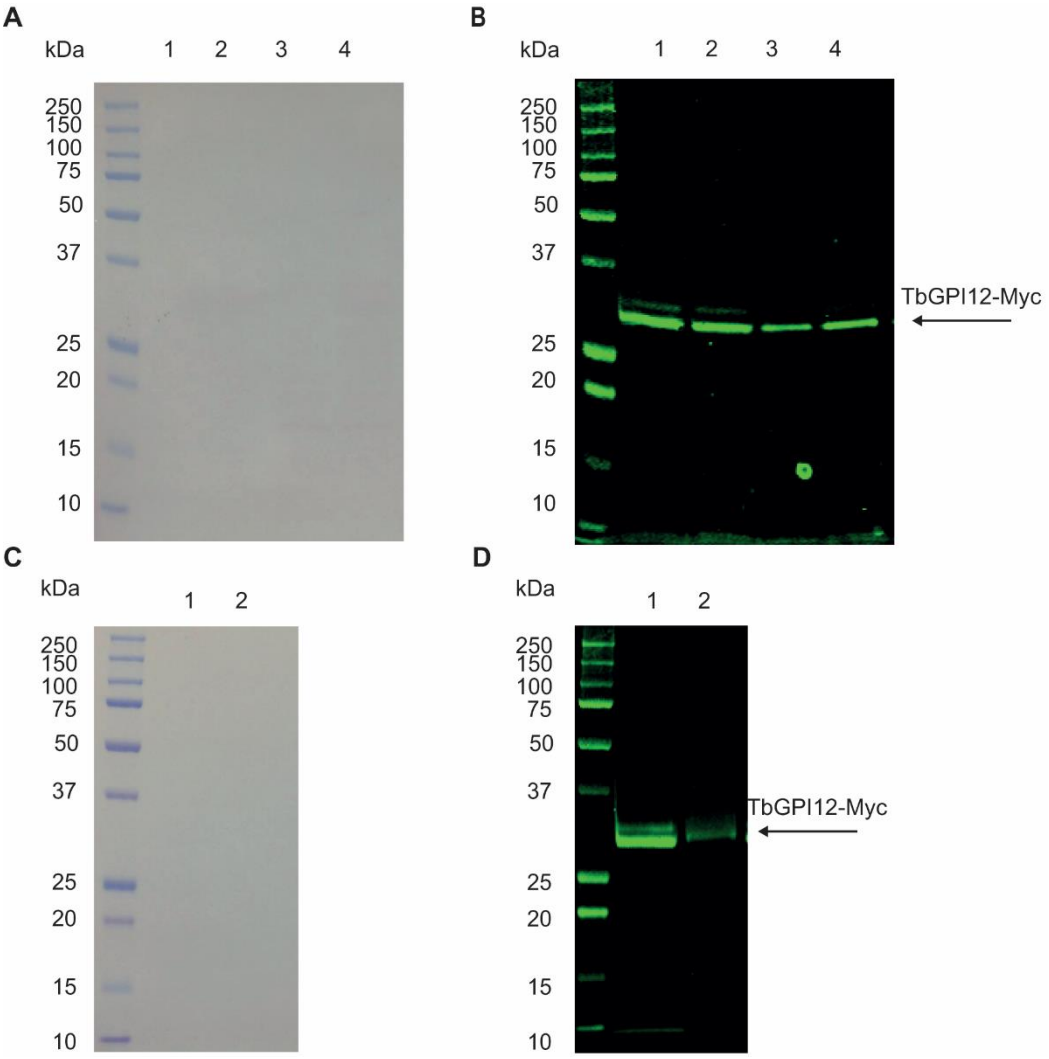

Figure 4.3 Introducing Myc-Trap™ agarose beads for IP and new c-Myc peptide as elution buffer.

A and B: TbGPI12-Myc PCF 0.5% digitonin lysates were sequentially immunoprecipitated twice using either Myc-Trap™ or home-made anti-Myc coupled magnetic beads. Proteins bound to the beads were eluted three times with pH 2.8 sodium citrate buffer and combined samples were analysed by SDS-PAGE followed by Ponceau staining and anti-Myc Western blot. Lanes 1 and 2: two sequential IPs with MycTrap™ beads; lanes 3 and 4: two sequential IPs with anti-Myc coupled magnetic beads. C and D: As panel A and B, except Myc-Trap ™ beads were eluted with new (Thermo) c-Myc peptide (lane 1) or pH 2.8 sodium citrate (lane 2).

4.1.5 Identification of protein complexes containing TbGPI12-Myc and of co-IP proteins by SILAC proteomics

To investigate whether TbGPI12-Myc forms a protein complex or complexes, TbGPI12-Myc protein was extracted and immunoprecipitated from the Tet-induced *TbGPI12* conditional null mutant cell line using the conditions described in Section

4.1.5. The eluted proteins were applied to a NativePAGE gel and detected by anti-Myc antibody. Wild type cells were included as negative control in the same experiment. Two bands were recognized in the TbGPI12-Myc samples; one fairly discrete at ~60 kDa, the other a broad smear running with and above the 146 kDa marker (Fig 4.4 A). The apparent molecular weights (MW) of these two bands are greater than the theoretical MW of TbGPI12-Myc (~25 kDa), indicating that TbGPI12 does form a complex with other proteins and/or itself. To further identify the proteins that co-IP with TbGPI12, stable isotope labelling in cell culture (SILAC) proteomics was performed as described in Materials and Methods. Briefly, procyclic wild type (PCF WT) and TbGPI12-Myc expressing parasites were grown under identical conditions for nine cell divisions, except that the TbGPI12-Myc cell line was grown in "heavy medium" containing stable isotope-labelled Lys and Arg (R6K4), whereas the PCF WT cells were grown in "light medium" containing unlabelled Lys and Arg (R0K0). The TbGPI12-Myc and the PCF WT cells were counted, mixed together in a 1:1 ratio and then harvested, washed and lysed in 0.5 % digitonin buffer. The IP was subsequently performed and the proteins eluted from the Myc-Trap^{TM} beads were processed to tryptic peptides for LC-MS/MS analysis. The data set from the experiment was used to search a *T. brucei* predicted protein database using MaxQuant software. Each protein was displayed on a plot of the Log10 transform of the Andromeda protein score calculated by MaxQuant (abbreviated as protein score hereafter) (y axis) and the Log2 value of the heavy to light isotope ratios (x axis) (Figure 4.4 B). All the proteins that have a heavy/light ratio greater than 4 were annotated in red, of which the bait protein TbGPI12-Myc had the third highest ratio. There were 13 other proteins identified that might potentially physically associate with TbGPI12-Myc (Table 4.1). The majority of these co-IP proteins have a relatively low protein ranking, suggesting that they are high abundance proteins which indicates that they are unlikely to be true partner proteins of the low abundance bait protein TbGPI12. Two of the hits, Tb927.11.6210 (lanosterol 14-alpha demethylase) and Tb927.10.7100 (delta-4 fatty acid desaturase), are relatively low abundance proteins that co-immunoprecipitated with TbGPI12. However, the associations between them need to be further validated.

Of all the proteins in the TbGPI12-Myc IP, TbMTI (TbGPI14) was notably absent suggesting that, contrary to our hypothesis, TbGPI14 may not be in a complex with TbGPI12. This may be due to the fact that TbGPI12 simply does not have physical association with TbGPI14, or that the association between TbGPI12 and TbGPI14 is real but only transient, such that the complex they form could not be obtained under the

IP conditions used. Thus, we next investigated cross-linking methodology which is suitable for stabilising protein complexes before IP experiments.

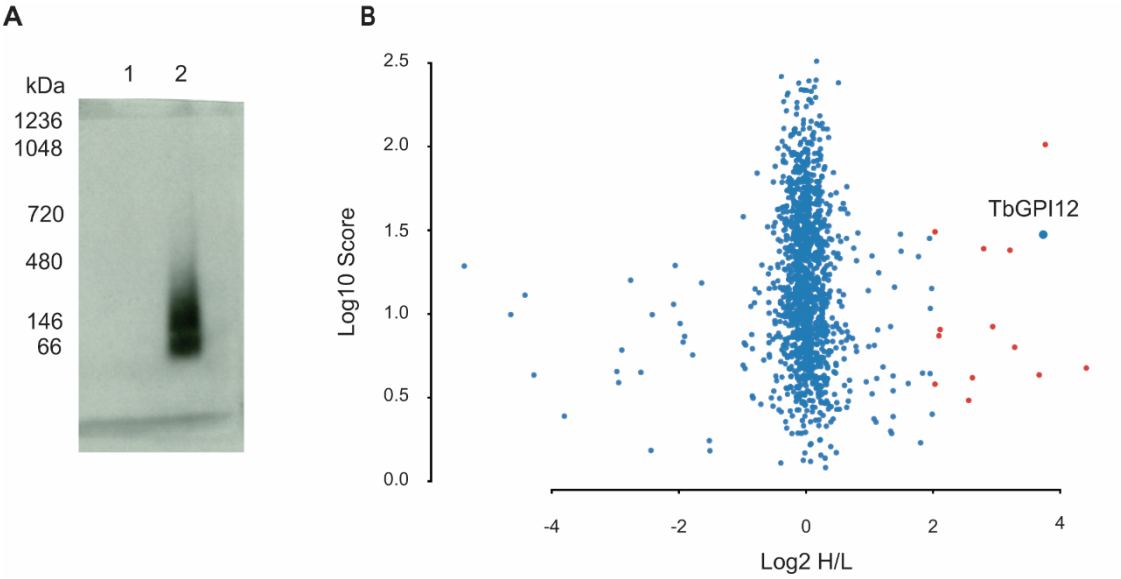

Figure 4.4 Identification of protein complexes containing TbGPI12-Myc and identification of potential co-IP proteins through SILAC proteomics.

A: NativePAGE Western blot of proteins immunoprecipitated from WT (lane 1) and TbGPI12-Myc expressing (lane 2) PCF cell lysates. B: Plot of TbGPI12-Myc/WT SILAC IP results. The plot shows the Log2 of the heavy-to-light isotope ratio (x axis) versus the Log10 of protein score (y axis). The TbGPI12-Myc (bait) protein is highlighted and proteins whose heavy/light ratio are greater than 2 are in red.

Table 4.1 List of co-IP proteins of TbGPI12 identified through SILAC proteomics.

T. brucei **gene (TriTrypDB gene ID)**	**Gene Description**	**Rank in TbGPI3-3Myc pull-downs[1]**	**Rank in total *T. brucei* cell Proteome**
Tb927.11.9400	Hypothetical protein, conserved	1	684
Tb927.5.1210	short-chain dehydrogenase, putative	2	250
Tb927.11.12080	N-acetyl-D-glucosaminylphosphatidylinositol de-N-acetylase ((TbGPI12)	3	5818
Tb927.11.6210	Lanosterol 14-alpha demethylase	4	7063
Tb927.10.4040	3-keto-dihydrosphingosine reductase	5	1667
Tb927.9.9630	TPR repeat/Tetratricopeptide repeat, putative	6	1474
Tb927.9.9710	Histidine phosphatase superfamily (branch 1), putative	7	2842
Tb927.6.4210	aldehyde dehydrogenase, putative	8	1732
Tb927.11.8380	NADPH--cytochrome P450 reductase, putative	9	1414

Tb927.10.7100	delta-4 fatty acid desaturase	10	6629
Tb927.11.12500	hypothetical protein, conserved	11	3314
Tb927.9.2050	membrane protein YIP1, putative	12	3082
Tb927.10.14510	root hair defective 3 GTP-binding protein (RHD3), putative	13	1089

4.2 Investigating co-IP proteins of TbGPI12-Myc using cross-linking proteomics.

4.2.1 Optimisation of cross-linking conditions for TbGPI12-Myc

Since both TbGPI12 and TbGPI14 are predicted to be transmembrane proteins, a cross linking methodology appropriate for studying membrane protein association was established (Fig 4.5A). Basically, PCF *T. brucei* cells expressing TbGPI12-Myc and PCF WT cells were subjected to hypotonic lysis and, after centrifugation, the supernatant containing the majority of soluble cytosolic proteins was removed. Subsequently, the cross-linking reaction was performed on the resuspended membrane pellets using the chemical crosslinker DTSSP, see (Section 3.7.2) for details of how it works. After quenching the reactions, the membrane proteins were solubilized with detergent 2% TX-100 and IP was performed, as described above. To achieve a favourable cross-linking reaction, different concentrations of DTSSP were first tested. As shown in (Fig 4.5 C), smeared bands that ran with higher apparent molecular weights than TbGPI12-Myc were detected in the cross-linked samples, and the signals of these bands became stronger as the concentration of DTSSP increased. Of these conditions, the 4 mM DTSSP condition was chosen for subsequent scaled up cross linking IP experiments and subsequent proteomics because: First, more TbGPI12-Myc was present in complexes using 4 mM DTSSP when compared to 1 and 2 mM DTSSP. Second, we considered that 8 mM DTSSP might lead to more non-specific cross-linking of proteins with TbGPI12-Myc.

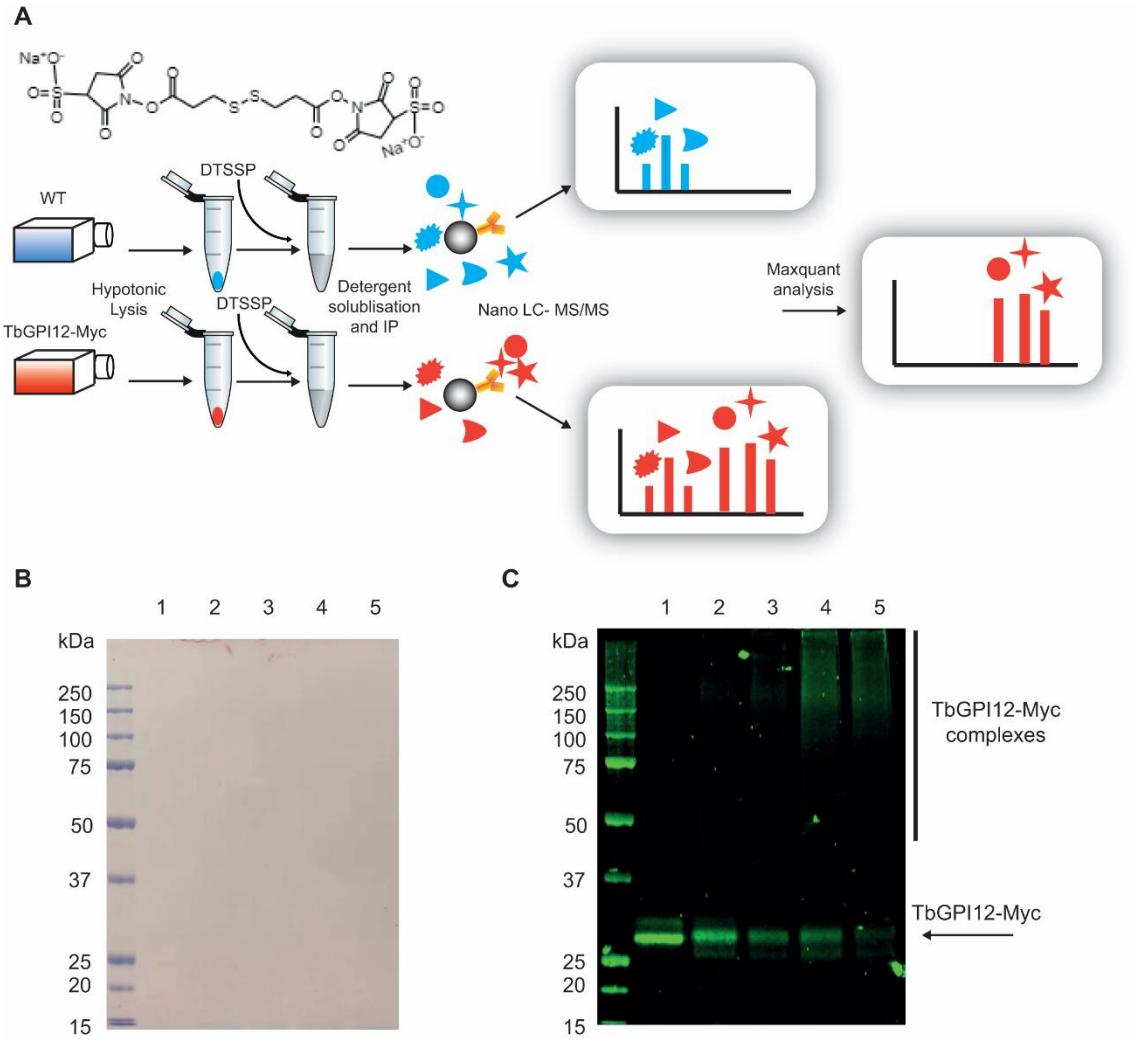

Figure 4.5 Summary of cross-linking proteomics experiment.

A: Scheme of cross-linking quantitative proteomics to identify proteins that specifically co-IP with TbGPI12-Myc. B and C : Optimisation of the DTSSP crosslinker concentration: TbGPI12-Myc expressing cells were lysed and the membrane fraction was collected and cross-linking reactions were performed using 0, 1, 2, 4 and 8 mM DTSSP (lanes 1-5, respectively). Membrane proteins were then solubilised in detergent and IPs were performed. Eluted proteins were separated on SDS PAGE. Ponceau staining (B) and anti-Myc Western blotting (C) were performed on the transferred membrane.

4.2.2 Identification of co-IP proteins of TbGPI12-Myc through cross linking proteomics

A scaled up IP experiment was then performed using the 4 mM DTSSP cross-linking condition and the proteins that were collected from the IP were applied to SDS-PAGE and gel regions that are outlined in the red (Fig 4.6 A) were cut and sent for label-free quantitative proteomics. The proteins identified from mass spectrometry were plotted by

the ratio of their Maxquant computed intensities detected in the TbGPI12-Myc IP sample (Bait) to the sum intensities (Sum) in both TbGPI12-Myc (bait) and PCF WT (control) IP samples (x axis) and the Log10 value of protein score (y axis) (Fig 4.6 B). This atypical x-axis value was adopted so that proteins undetected in the control are not lost in the analysis, as can happen when enrichments are "infinite" or "infinitesimal", i.e., when proteins appear exclusively in only one sample of a pairwise comparison. Thus, in this experiment, TbGPI12 and the proteins that specifically associated with it should have a ratio very close to 1. Also, the quality of peptide identification by mass spectrometry was taken into consideration by applying a cut off of proteins whose scores were less than 15. Using this calculation, 16 protein groups were identified with a ratio of 1 including the bait protein TbGPI12-Myc. The other 15 proteins are potential co-IP proteins of TbGPI12 (Table 4.2), however, TbGPI14 was not among these.

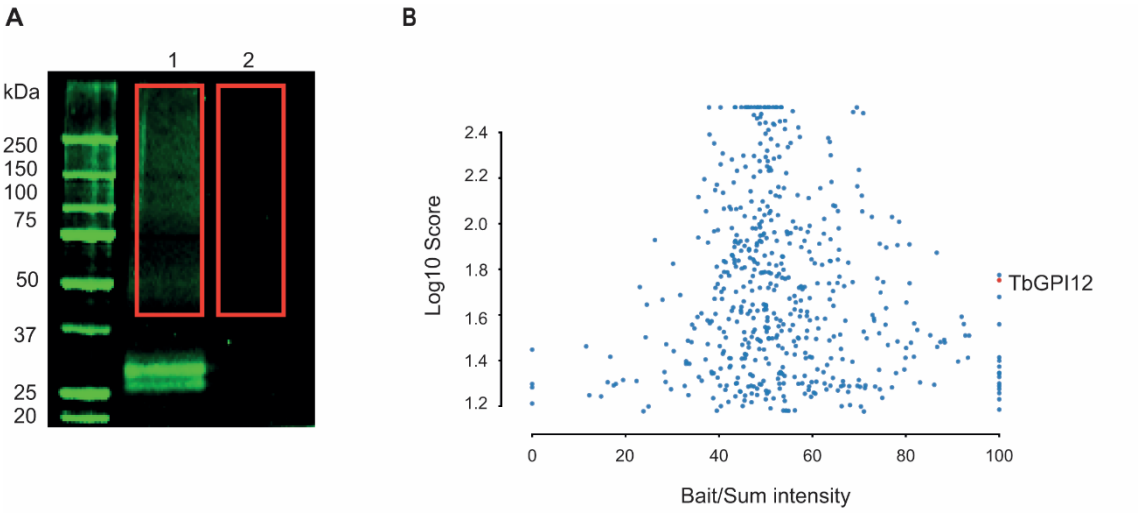

Figure 4.6 Identification of TbGPI12-Myc co-IP proteins following membrane protein cross linking.

A: Cross-linking experiments were performed on TbGPI12-Myc expressing and WT PCF washed cell membranes as described above using 4 mM DTSSP. An anti-Myc Western blot was performed on the proteins that were eluted after IP. A gel that was run in parallel was cut as indicated by the red rectangles and the gel slices sent for proteomics. Each gel slice was divided into four sub-slices which were processed to tryptic peptides and subjected to LC-MS/MS separately. Concatenated data from the four sub-slices were processed by Maxquant. B: Plot of label free quantification proteomics data obtained from the regions shown in panel A. X axis: ratio of protein intensities detected in the TbGPI12-Myc IP sample compared with the sum of

their intensities detected in the TbGPI12-Myc and PCF WT samples. Y axis: log 10 value of the Andromeda protein score calculated by Maxquant.

Table 4.2 List of co-IP proteins of TbGPI12 identified by cross-linking proteomics.

T. brucei **gene (TriTrypDB gene ID)**	**Gene Description**	**Rank in total *T. brucei* cell Proteome**
Tb927.11.15990	Hypothetical protein, conserved	848
Tb927.11.12080	N-acetyl-D-glucosaminylphosphatidylinositol de-N-acetylase ((TbGPI12)	5818
Tb927.5.1210	Nucleoporin NUP152	1358
Tb927.11.2950	Nucleoporin NUP89	1576
Tb927.10.12240	3-ketoacyl-CoA reductase, putative	2486
Tb927.10.14510	root hair defective 3 GTP-binding protein (RHD3), putative	1089
Tb927.11.7060	acidocalcisomal pyrophosphatase	5598
Tb927.6.3570	tubulin-tyrosine ligase-like protein, conserved	6779
Tb927.11.13250	eukaryotic translation initiation factor 2 gamma, putative	637
Tb927.11.7560	Component of motile flagella 15	433
Tb927.11.9130	Replication factor A protein 1	956
Tb927.8.2630	kinesin-C	837
Tb927.4.2080	Coiled-coil and C2 domain-containing protein	256
Tb927.7.3510	LSU ribosomal protein, mitochondrial, putative	6192
Tb927.7.1070	KREX1	5995
Tb927.7.5160	deoxyuridine triphosphatase, putative	1769

4.3 Co-IP proteins in common between with the SILAC and cross-linking proteomics datasets.

The two lists of proteins that co-IP with TbGPI12-Myc were generated through two different types of IP proteomics experiments. The proteins identified from these two results are overlapped (Fig 4.7) considering the following criteria: For the SILAC proteomics data, the protein heavy/light ratios should be >4. For the cross-linking proteomics data, only proteins whose Bait/Sum intensity ratio are 1 are considered. For both experiments, the proteins whose protein scores were less than 15 were removed. Using these stringent criteria, there are two proteins in common between these two datasets, and these are ranked by their heavy/light isotope ratio from the SILAC

experiment (Table 4.3). The first of these is the TbGPI12-Myc bait protein, as expected. Since TbMTI (TbGPI14) did not appear in either dataset, or in in preliminary co-IP experiment, see (Section 7.1), we conclude that there is no evidence to support a physical association between TbGPI12 and TbGPI14 from these sensitive proteomics approaches. The most likely candidate that may have physical association with TbGPI12-Myc is Tb927.10.14510. According to TritrypDB, this gene encodes a putative root hair defective 3 GTP-binding protein (RHD3). It contains a putative transmembrane domain near its C-terminus and is localised in the ER (Dean *et al*., 2017). According to motif search, it has an RHD3 (PF05879 Pfam) motif which is thought to be a motif for a family of a novel class of GTP-binding proteins. Previous studies on the function(s) of the proteins containing this motif suggests that they are required for regulated cell enlargement (Wang *et al.*, 1997) or involved in membrane trafficking (Brands and Ho, 2002), neither of which directly supports a potential function in the GPI anchor biosynthesis pathway. It is possible that Tb927.10.14510 could be a transiently associating regulatory protein for TbGPI12, however, more specific functional studies would need to be carried out for further characterisation through reverse pull down of epitope tagged of Tb927.10.14510.

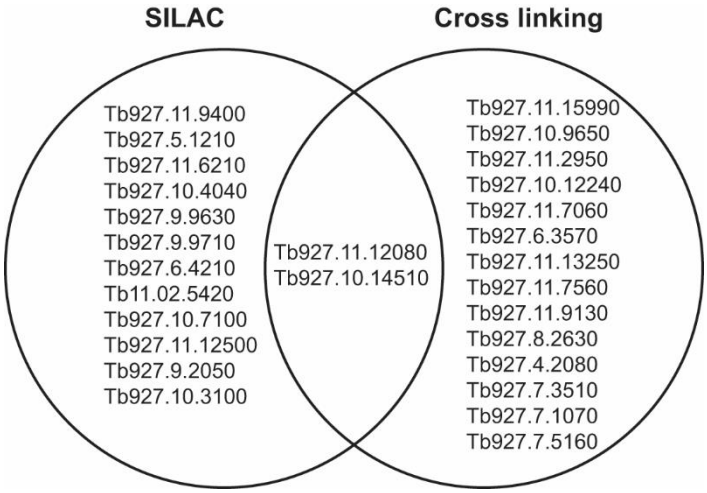

Figure 4.7 Overlapping of TbGPI12-Myc associated proteins identified by SILAC and cross-linking IP proteomics.

Table 4.3 List of proteins that co-IP with TbGPI12-Myc identified by both SILAC and cross-linking proteomics.

Gene ID	Gene description	SILAC ratio (H/L)
Tb927.11.12080	N-acetyl-D-glucosaminylphosphatidylinositol de-N-acetylase	13.297

Tb927.10.14510	root hair defective 3 GTP-binding protein (RHD3), putative	4.0855

5 Results II- Studies on the *Trypanosoma brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase (GPI GnT) complex

5.1 Identification of putative *T. brucei* UDP-GlcNAc : PI α 1-6 GlcNAc-transferase complex components.

Conventional BLASTp searches with default settings (Acland *et al.*, 2012) were sufficient to identify *T. brucei* homologues of PIGA(GPI3), PIGC(GPI2), PIGP(GPI19), PIGQ(GPI1) and DPM2. However, the results for PIGH(GPI15) and PIGY(ERI1) were equivocal so a Domain Enhanced Lookup Time Accelerated BLAST (Boratyn *et al.*, 2012) using a PAM250 matrix was applied to find the corresponding *T. brucei* homologues (Table 5.1).

Table 5.1 Genes encoding UDP-GlcNAc : PI α1-6 GlcNAc-transferase (GPI GnT) complex subunits in mammalian cells, yeast and *T. brucei*.

Mammalian gene	Yeast gene	*T. brucei* gene (TriTrypDB gene ID)	Calculated MW of *T. brucei* gene product	Rank in TbGPI3-3Myc pull-downs[1]	Rank in total *T. brucei* cell Proteome[2]
PIGA	*GPI3*	*TbGPI3* (Tb927.2.1780)	51.6 kDa	1	6475
PIGH	*GPI15*	*TbGPI15* (Tb927.5.3680)	28.8 kDa	2	7038
PIGP	*GPI19*	*TbGPI19* (Tb927.10.10110)	16.5 kDa	3	6185
PIGC	*GPI2*	*TbGPI2* (Tb927.10.6140)	38.5 kDa	4	5285
PIGQ	*GPI1*	*TbGPI1* (Tb927.3.4570)	81.5 kDa	5	6797
		TbArv1 (Tb927.3.2480)	28.7 kDa	6	7150
PIGY	*ERI1*	*TbERI1* (Tb927.4.780)	10.1 kDa	7	7150
		TbUbCE (Tb927.2.2460)	22.6 kDa	8	2538
DPM2	-	*TbDPM2* (Tb927.9.6440)[3]	9.3 kDa	not present	7150

[1]Taken from ((Ji *et al*., 2021) S2 File).

[2]Taken from ((Ji *et al*., 2021) S1 File) where proteins are ranked from 1 (most abundant) to 7149 (least abundant) and undetected proteins are ranked 7150.

[3]Probable mis-annotation, see text.

5.2 *In situ* epitope tagging of TbGPI3.

To investigate whether a multi-subunit UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex might exist in *T. brucei* we selected TbGPI3, which encodes a 455 amino acid protein with two predicted transmembrane domains, one near its N-terminus and one

near its C-terminus (Käll, Krogh and Sonnhammer, 2004), for epitope tagging. We chose this PIGA(GPI3) homologue as the bait protein because PIGA has been shown to have either direct or indirect interactions with all other subunits in the mammalian UDP-GlcNAc : PI α 1-6 GlcNAc-transferase complex (Kinoshita, 2020). Alignment of putative TbGPI3, yeast GPI3 and PIGA protein sequences show that the *T. brucei* sequence has 43.9% and 50.8% sequence identity with the yeast and human sequences, respectively (Fig 5.1).

In situ tagging of the TbGPI3 gene was achieved by transfecting BSF *T. brucei* with PCR products amplified from the pMOTag43M plasmid (Oberholzer *et al.*, 2006) with primers OL1 and OL2, (Fig 5.2A). Transfected cells were selected using hygromycin and subsequently cloned by limit-dilution. Lysates of four separate clones were subjected to anti-Myc Western blotting (Figs 1B and 1C). *In situ* tagged TbGPI3-3Myc protein was detected in all four clones at an apparent molecular weight of ~47 kDa, somewhat lower than the predicted molecular weight of 55 kDa (Fig 5.2C, lanes 1-4). The specificity of this staining, although weak, is illustrated by the absence of comparable staining for the un-transfected cell sample (Fig 5.2 C, lane 5). The weakness of the TbGPI3-3Myc staining with anti-Myc is a function of the extremely low abundance of TbGPI3 in the total cell proteome, where it ranks 6475th out of 7149 detectable protein groups (Table 5.1), and the limitations of protein loading for BSF cell lysates on SDS-PAGE caused by the abundance of VSG and tubulin in these cells from their surface coat and subpellicular cytoskeleton, respectively. These can be seen running above and below the 50 kDa marker in (Fig 5.2B).

```
GPI3            MG-----------------------------FNIAMLCDFFYPQLGGVE
PIGA            MACRGGAGNGHRASATLSRVSPGSLYTCRTRTHNICMVSDFFYPNMGGVE
TbGPI3          MR-----------------------------RHRVAVVSDFFYPGFGGVE
                *                               ..:.::.***** :****

GPI3            FHIYHLSQKLIDLGHSVVIITHAYKDRVGVRHLTNGLKVYHVPFFVIF--
PIGA            SHIYQLSQCLIERGHKVIIVTHAYGNRKGIRYLTSGLKVYYLPLKVMY--
TbGPI3          VHIYSLGQCLMRRGHKVIVITRAYGDTCGVRYLTNGMKVYYLPLMAVKLP
                 *** *.* *:  **.*:::*:** :  *:*:**.*:***::*: .:

GPI3            -RETTFPTVFSTFPIIRNILLREQIQIVHSHGSASTFAHEGILHANTMGL
PIGA            -NQSTATTLFHSLPLLRYIFVRERVTIIHSHSSFSAMAHDALFHAKTMGL
TbGPI3          AGSVTLPTMYLTFATMRSIFIRERITVVHGHQNTSNLCHEALFHAGTLGL
                  . * .*:: ::. :* *::**:: ::*.* . * :.*:.::** *:**

GPI3            RTVFTDHSLYGFNNLTSIWVNKLLTFTLTNIDRVICVSNTCKENMIVRTE
PIGA            QTVFTDHSLFGFADVSSVLTNKLLTVSLCDTNHIICVSYTSKENTVLRAA
TbGPI3          KTCFTDHSLFGFADVSSIHINKVCEWSLRNVDQVICVSNTSRENTVLRAK
                :* ******:** :::*:  **:   :* : :::**** *.:** ::*:

GPI3            LSPDIISVIPNAVVSEDFKPRDPTGG--TKRKQSRDKIVIVVIGRLFPNK
PIGA            LNPEIVSVIPNAVDPTDFTPDPF---------RRHDSITIVVVSRLVYRK
TbGPI3          IDPQRVSVIPNATDCSFFTPPDDMKYKSWASKVENEGLTIVVIGRLVYRK
                :.*: :******.    *.*             .: :.***:.**. .*

GPI3            GSDLLTRIIPKVCSSHEDVEFIVAGDGPKFIDFQQMIESHRLQKRVQLLG
PIGA            GIDLLSGIIPELCQKYPDLNFIIGGEGPKRIILEEVRERYQLHDRVRLLG
TbGPI3          GSDLFVDVIPEICKRHPNIRWIVGGDGPRRSQFQQMIERHDLMDRVKMLG
                * **:  :**::*. : ::.:*:.*:**:   :::: * : * .**::**

GPI3            SVPHEKVRDVLCQGDIYLHASLTEAFGTILVEAASCNLLIVTTQVGGIPE
PIGA            ALEHKDVRNVLVQGHIFLNTSLTEAFCMAIVEAASCGLQVVSTRVGGIPE
TbGPI3          SLPHSGVRNVLIQGQIFLNCSLTEAFCIALIEAASCGLLCVSTRVGGVPE
                :: *. **:** **.*:*: ******   ::*****.*  *:*:***:**

GPI3            VLPNEMTVYAEQTSVSDLVQATNKAINIIRSK-ALDTSSFHDSVSKMYDW
PIGA            VLPENLIILCEP-SVKSLCEGLEKAIFQLKSGTLPAPENIHNIVKTFYTW
TbGPI3          VLPPNMLLLAEP-DPSSIITTLEEAIASVP---YISPWELHDNVRRFYRW
                *** :: : .*  . ..:    ::**  :       . .:*: *  :* *

GPI3            MDVAKRTVEIYTNISSTSSADDKDWMKMVANLYKRDGIWAKHLYLLCGIV
PIGA            RNVAERTEKVYDRVSVEAVLPMDK---RLDRLISHCGPVTGYIFALLAVF
TbGPI3          DWVAERTERVYDKIMCTKSPSLYE---RLMNYAS-VGCVYGVICWLLCIG
                  **:** .:* .:         .    : .  .  *     :  *  :

GPI3            EYMLFFLLEWLYPRDEIDLAPKW--PKKTVSNETKEARE----------T
PIGA            NFLFLIFLRWMTPDSIIDVAIDATGPRGAWTNNYSHSKRGGENNEISETR
TbGPI3          DWLMLTFLEFWFPSELIDIAPDF--PLELYSRNREKLQVVGSPS------
                ::::: :*.:  * . **:* .   *    :.: .. :
```

Figure 5.1 Alignment of PIGA and TbGPI3 was generated by the Tcoffee program.

* = identical residues, : = conserved residues. GPI3, PIGA and TbGPI3 sequences are from Uniprot; GPI3 (P32363 GPI3_YEAST) PIGA (P37287 (PIGA_HUMAN)); TbGPI3 (Q585S5 (Q585S5_TRYB2)).

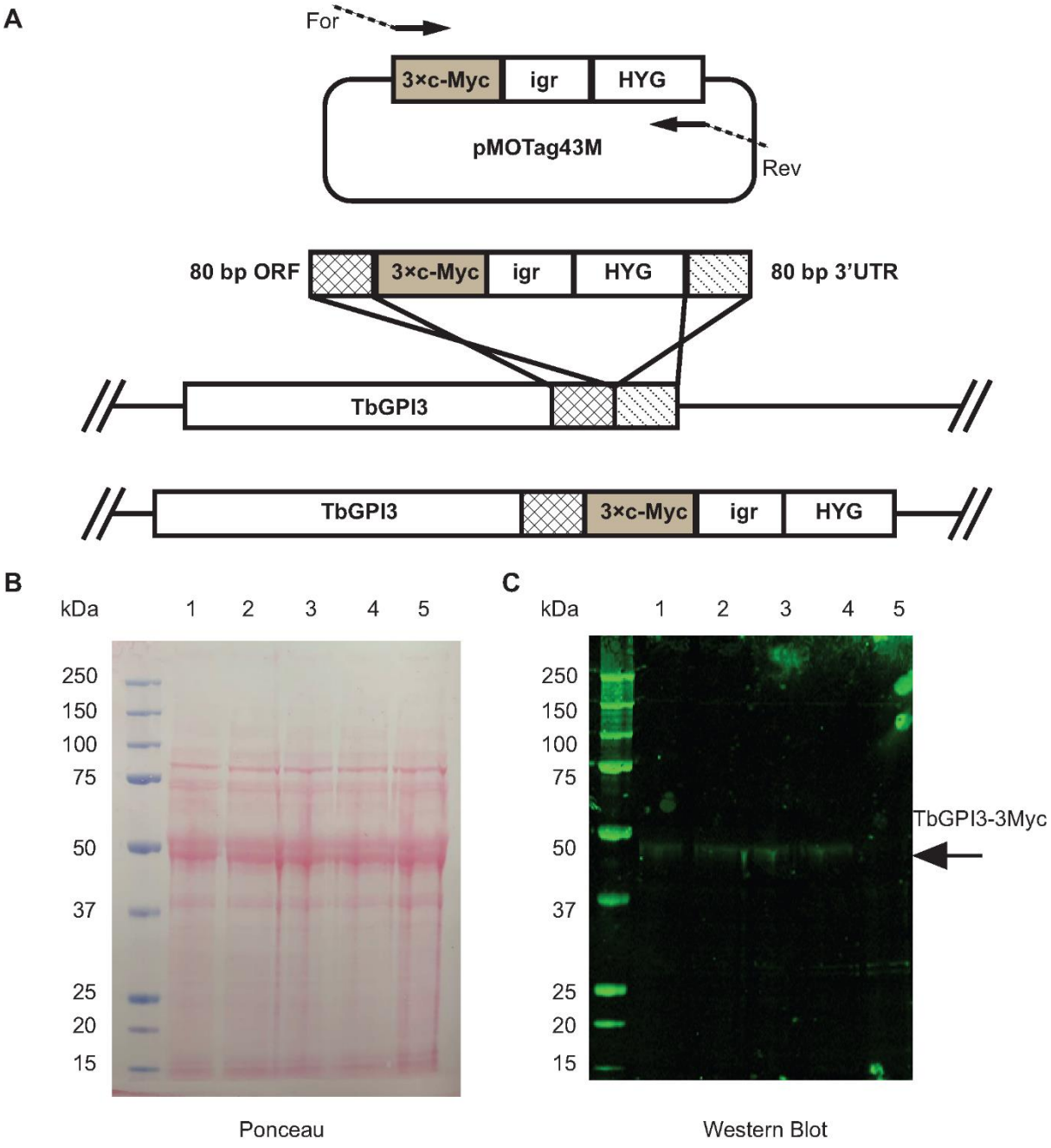

Figure 5.2 *In situ* C-terminal tagging of TbGPI3 with 3Myc.

A: Map of plasmid pMOTag43M (Oberholzer *et al.*, 2006) used for the in situ tagging of TbGPI3, and a scheme of how the PCR product generated with the indicated forward (For) and reverse (Rev) primers inserts into the 3'-end of the TbGPI3 ORF (checked box) and 3'-UTR (striped box) in the parasite genome to effect in-situ tagging. HYG = hygromycin phosphotransferase selectable marker; igr = α-β tubulin intergenic region. B: Ponceau staining of denaturing SDS-PAGE Western blot shows similar loading and transfer of lysates (corresponding to 5×10^6 cells) from four in-situ tagged clones (lanes 1-4) and wild type cells

(lane 5). C: An identical blot was probed with anti-Myc antibody. TbGPI-3Myc is indicated by the arrow. The positions of molecular weight markers are indicated on the left of (B) and (C).

5.3 Solubilisation and native-PAGE of TbGPI3-3Myc.

The analysis of epitope-tagged membrane bound multiprotein complexes requires detergent extraction and anti-epitope pull-down under conditions that preserve intermolecular interactions within those complexes. To investigate detergent extraction conditions, TbGPI-3Myc expressing cells were cultured and lysed with 0.5% digitonin, 1% digitonin, 1% TX-100, 1% NOG and 1% DM and centrifuged. The solubilised proteins in the supernatants from these treatments, along with a 1% TX-100 extract of wild type cells, were immunoprecipitated with Myc-Trap^{TM} agarose beads that were washed three times and finally eluted with synthetic c-Myc peptide. The proteins in the eluates were separated by denaturing SDS-PAGE and by native PAGE (Schägger *et al*., 1991) and analysed by anti-Myc Western blot (Figs 5.3A and 5.3B, respectively). All detergents, apart from DM, extracted the TbGPI3-3Myc protein (Fig 5.3A). Of these conditions, 1% TX-100 gave the highest efficiency of extraction but duplicate samples analysed by native PAGE and anti-Myc Western blot showed that digitonin best preserved a TbGPI3-3Myc-containing complex with a native apparent molecular weight of ~240 kDa and that 0.5% digitonin gave a higher-yield of the complex than 1 % digitonin (Fig 5.3B). The reason for not detecting any clear complexes in other conditions may be due to the ~240 kDa complex falling apart into multiple sub-complexes below the limits of detection. In all of these experiments, we adopted a pull-down (from 2 x 10^8 cell equivalents) prior to immunodetection approach because of the aforementioned low abundance of TbGPI3 and its expected partner proteins (Table 1.1) and the correspondingly weak anti-Myc signals obtainable from 5 x 10^6 cell equivalents (Fig 5.3C).

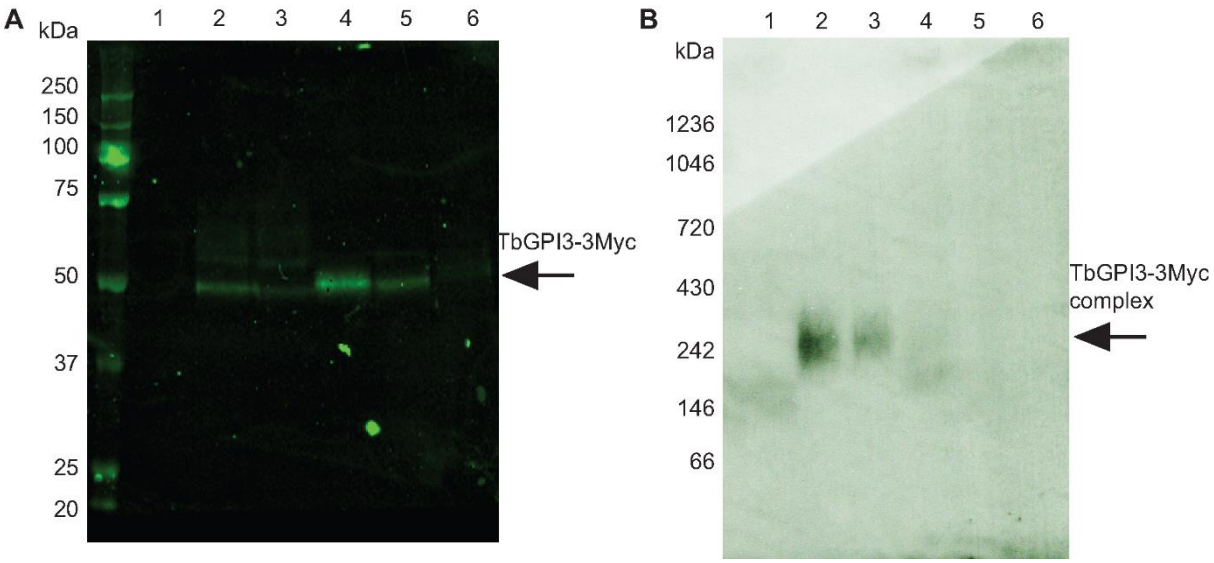

Figure 5.3 TbGPI3-3Myc is present in complexes in BSF *T. brucei*.

A: Aliquots of 2×10^8 cells were harvested and lysed in lysis buffer containing different detergents to assess TbGPI3-3Myc solubilisation. After immunoprecipitation of the supernatants with anti-Myc agarose beads, proteins were eluted with 0.5 mg/mL c-Myc peptide solution and aliquots were subjected to SDS-PAGE followed by anti-Myc Western blotting. B: Identical samples were also separated by native-PAGE and subjected to anti-Myc Western blotting. In both cases, lane 1 corresponds to wild type cells lysed with 1% TX-100 as a negative control and lanes 2-6 correspond to TbGPI3-3Myc clone1 lysed with 0.5% digitonin, 1% digitonin, 1% TX-100, 1% NOG or 1% DM, respectively.

5.4 Identification of *T. brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex components by quantitative proteomics.

Having established detergent solubilisation conditions that retained TbGPI13-3Myc in a complex, we performed label-free quantitative proteomics on Myc-Trap™ pull downs to identify the components within the complex. For this experiment, BSF WT and TbGPI13-3Myc expressing parasites were grown under identical conditions and the same numbers of cells were harvested and lysed in 0.5 % digitonin lysis buffer. Immunoprecipitation was performed using Myc-Trap™ beads and the proteins eluted from these two samples with SDS sample buffer were processed to tryptic peptides for LC-MS/MS analysis (Fig 5.4A). The experiments were performed in biological triplicates and the data were analysed using MaxQuant software and a newly developed data analysis method written in Python called ProtRank (Medo, Aebersold and Medová, 2019), see Materials and Methods. The protein groups identified ((Ji *et al*., 2021) S2 File) were displayed on a plot of the minus log10 value of their False Discovery Rate

(y-axis) and enrichment rank (x-axis) between the bait versus control samples (Fig 5.4B). As expected, the bait protein TbGPI3-3Myc was the most highly enriched protein and its putative partner proteins TbGPI15, TbGPI19, TbGPI2, TbGPI1 and TbERI1 were also significantly enriched. Notably, TbDPM2 (dolichol-phosphate-mannose synthetase 2) was not detected. However, although TbDPM2 is annotated in the TriTrypDB database it should be noted that, like yeast, *T. brucei* makes a single-chain dolichol-phospho-mannose synthetase (DPM1) (Mazhari-Tabrizi *et al.*, 1996) rather than a trimeric enzyme made of a soluble catalytic DPM1 subunit associated with small transmembrane DPM2 and DPM3 subunits, as found in mammalian cells. For these reasons, we feel that the absence of DPM protein components in the *T. brucei* complex is to be expected.

Interestingly, an Arv1-like protein (hereon referred to as TbArv1, Tb927.3.2480) and a putative ubiquitin-conjugating enzyme E2 (UbCE, Tb927.2.2460) were also co-immunoprecipitated with TbGPI3 (Fig 5.4C). The data were also processed in a different way (see Materials and Methods) that plots the experimental rank (x-axis) against the rank order of estimated abundances of the protein groups in the total cell proteome, generated from data in (Tinti, Maria Lucia S Güther, *et al.*, 2019), on the y-axis (Fig 5.4C). In this plot, the very low abundance TbArv1 (undetectable in the total cell proteome) clusters better with the canonical and similarly low abundance UDP-GlcNAc : PI α1-6 GlcNAc-transferase subunits. By contrast, although UbCE is clearly highly-enriched in the pull-down it is a much more abundant protein, suggesting that only some fraction of it may be associated with the complex.

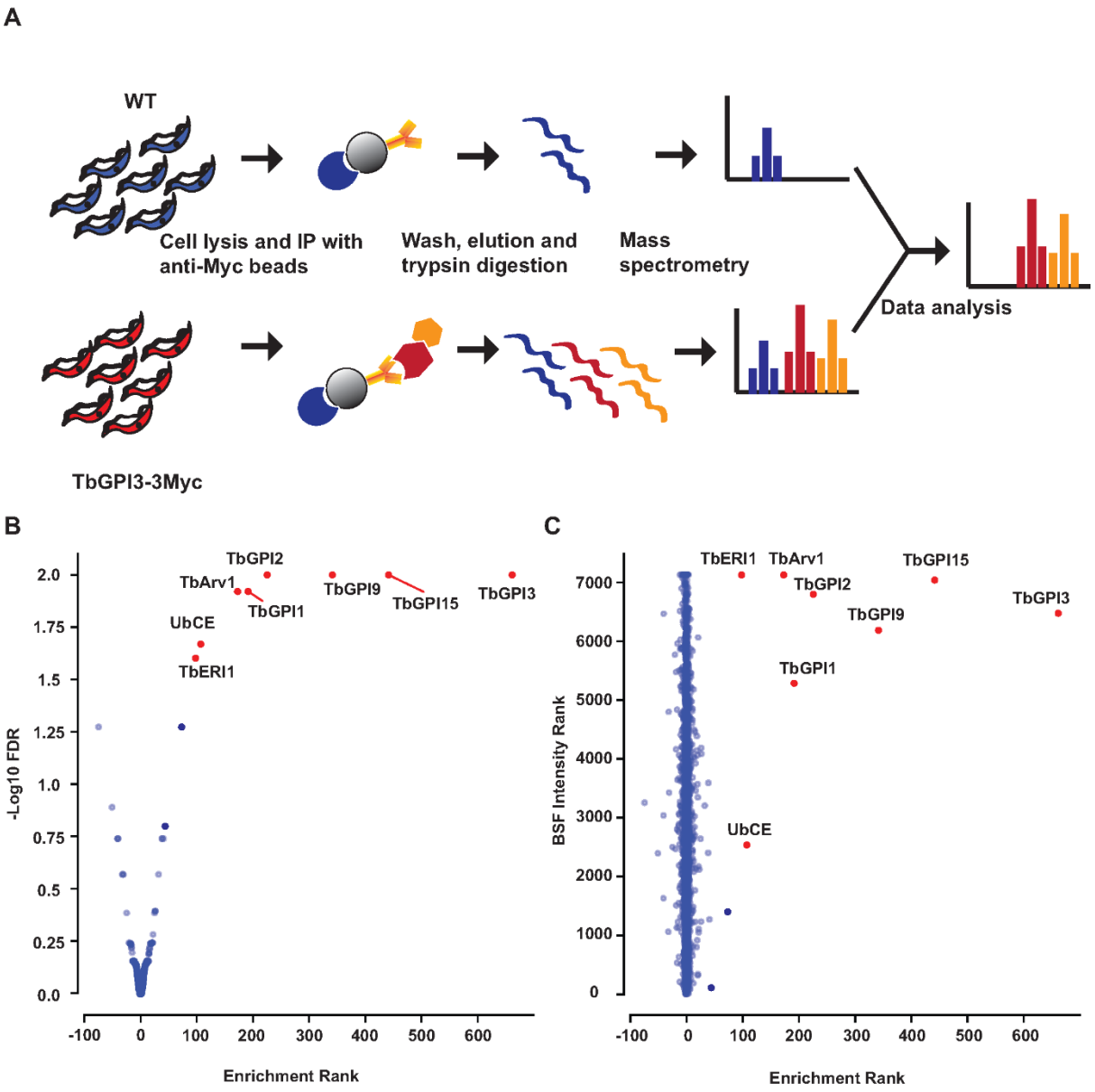

Figure 5.4 Identification of UDP-GlcNAc : PI α1-6 GlcNAc-transferase subunits by immunoprecipitation of TbGPI3-3Myc from BSF *T. brucei* digitonin lysates.

A: Scheme of the label-free proteomics approach to identify TbGPI13-3Myc binding partners. BSF WT and TbGPI13-3Myc expressing cell lines were cultured, harvested and lysed in 0.5% digitonin lysis buffer. Identical quantities of the supernatants were subjected to anti-Myc agarose bead immunoprecipitation and the bound proteins were eluted from the beads with SDS sample buffer. The eluted proteins were reduced, alkylated and digested with trypsin and the resulting peptides analysed by LC-MS/MS. B: Volcano plot comparing protein groups present in the anti-cMyc immunoprecipitates from TbGPI3-3Myc expressing cell lysates versus WT cell lysates. Mean values (from biological triplicate experiments) for each protein group (dots) are plotted according to their minus log10 False Discovery Rate values (y-axis), calculated by MaxQuant, and the enrichment rank (x-axis). The enrichment rank was computed with the ProtRank algorithm using the iBAQ values calculated by MaxQuant. The higher the rank value on the x-axis, the higher the abundance in the TbGPI3-3Myc samples. The putative subunits of

UDP-GlcNAc : PI α1-6 GlcNAc-transferase in *T. brucei* are highlighted in red and annotated with their corresponding names (Table 5.1). C: Relative intensity plot using a new algorithm. The same data as (B) are plotted with a different y-axis, whereby each protein group is assigned an intensity rank from the most abundant protein group (1) to least abundant protein groups (7149) based on their summed eXtracted Ion Currents (XICs) for the total BSF proteome. (Details of the mass spectrometry and data analysis are provided in in Materials and Methods.)

5.5 Investigating co-IP proteins of TbGPI3-3Myc using cross-linking proteomics.

5.5.1 Optimisation of cross-linking conditions for TbGPI13-3Myc

To investigate whether there might be more novel subunits in the GPI GnT complex in *T. brucei*, in particular possible soluble partner proteins, the cross-linking proteomics methodology established for TbGPI12 (section 4.2) was adapted and applied to TbGPI3. First, whole hypotonic cell lysate (membranes and cytosol), and the membrane fraction collected by centrifugation therefrom, were incubated with 8 mM DTSSP. The TbGPI3-3Myc protein was present in high apparent MW complexes in both samples (Fig 5.5A, lanes 2 and 4) as compared to the non DTSSP-treated controls (lanes 1 and 3). However, more TbGPI-3Myc appeared in the whole cell lysate sample (lane 2). This suggested that using whole cell lysate for the cross-linking reaction, rather than the membrane fraction, might be advantageous in identifying components of cross-linked TbGPI-3Myc-containing complexes.

Next, the concentration of DTSSP cross-linker was optimised. Under all conditions tested (1, 2, 4 and 8 mM DTSSP), high apparent MW protein complexes containing TbGPI3-3Myc were detected, with no significant differences between these samples (Fig 5.5B). Therefore, a concentration of 1mM DTSSP was chosen for subsequent experiments as a higher concentration crosslinker might lead to the identification of non-specific cross-linked partner proteins of TbGPI3-3Myc by proteomics.

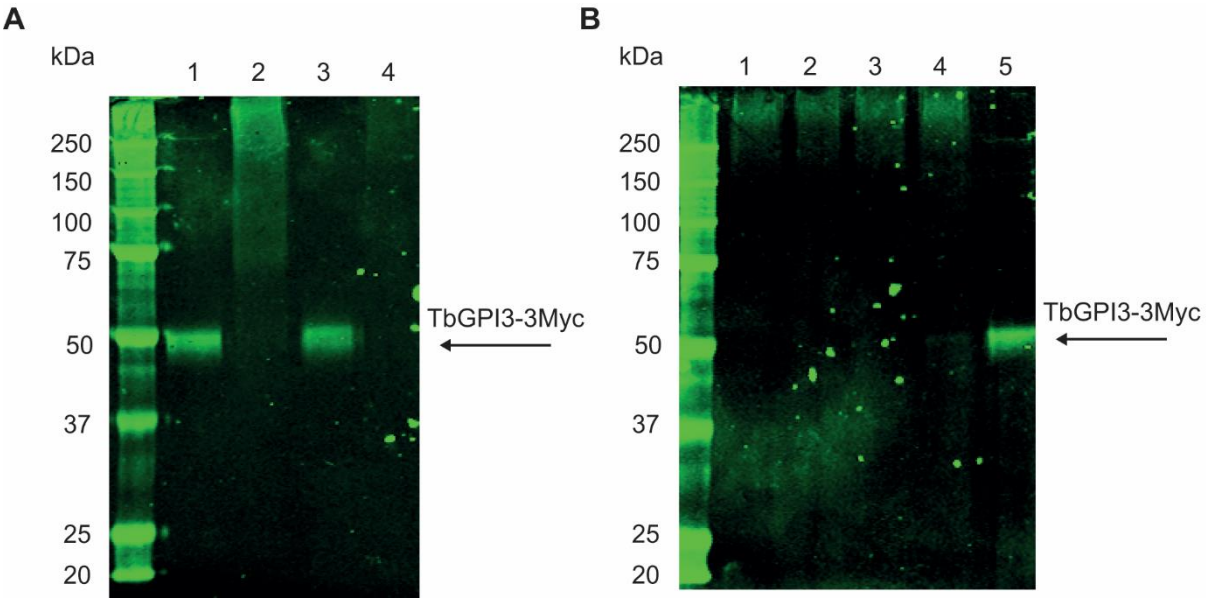

Figure 5.5 Optimisation of cross-linking reactions for TbGPI3-3Myc.

A: Comparison of two different sample preparations for cross-linking reactions. Lane 1 and 2: TbGPI3-3Myc expressing cells were hypotonic lysed and cross-linking reactions were performed to the whole cell lysate with 0 and 8 mM DTSSP. Lane 3 and 4: The membrane fraction of the cell lysates of the TbGPI3-3Myc expressing cell were collected and cross-linking reactions were performed with 0 and 8 mM DTSSP. Proteins were then solubilised with 2% TX-100 and IPs were performed. Eluted proteins were analysed by SDS-PAGE following by anti-Myc Western blot. B: Optimisation of the DTSSP crosslinker concentration: TbGPI3-3Myc expressing cells were lysed and cross-linking reactions were performed using 8, 4, 2, 1 and 0 mM DTSSP (lanes 1-5, respectively). IPs were performed on the 2% TX-100 solubilised proteins and an anti-Myc Western blot was performed subsequently on the eluted proteins.

5.5.2 Identification of co-IP proteins of TbGPI3-3Myc through cross linking proteomics

A scaled-up cross-linking reaction was then set up for the whole cell lysate of TbGPI3-3Myc expressing cells using 1 mM DTSSP followed by 2% TX-100 solubilisation and anti-Myc immunoprecipitation. The eluted proteins were separated by SDS-PAGE and the gel regions that contained TbGPI3-3Myc complexes, and the corresponding BSF WT control counterparts (Fig 5.6A, lane 1 and 2, outlined in red) were cut and send for label free quantitative proteomics identification. The proteins identified were plotted as the log2 values of the ratios (TbGPI3-3Myc IP sample (bait) to the BSF WT (control) IP sample) of their Maxquant computed intensities (x axis) and the Log10 value of sum of their intensities detected in both samples (y axis) (Fig 5.6B). A protein cut-off score of 5

was also applied, to control for protein identification quality by mass spectrometry, and the proteins with log2 bait/control values >2 are listed in (Table 5.2).

Only TbGPI3 out of eight putative GPI GnT subunits (Table 5.1) was enriched and identified by this approach, and the quality of the protein identification by mass spectrometry was not ideal with calculated protein scores less than 15. Of the other proteins: Tb927.6.3800 is unlikely to be a specific partner protein for TbGPI3-3Myc as it thought to be a mitochondrial heat shock 70 (HSP70) protein. It is notable that it is a very abundant cellular protein according to its ranking from the total *T. brucei* cell proteomics data. Tb927.9.14280 is a 187 kDa cytosolic protein and contains two protein domains which are a concanavalin A-like lectin domain (Con A-like domain) (SSF49899) and an armadillo (ARM) repeat superfamily domain (SSF48371) annotated in the SUPERFAMILY dataset (Wilson *et al.*, 2007). Proteins containing Con A-like domains are predicted to be involved in cell recognition and adhesion. However, Con-A itself is an α-Man-specific lectin, and GPI intermediates (Man_2GlcN-PI and beyond) are able to bind to it (Vidugiriene and Menon, 1994). Thus, it is conceivable that Tb927.9.14280 could bind to and possibly sense GPI intermediate levels and regulate the first step of the pathway, however, this is highly speculative. While there are several different classes of function for proteins containing ARM repeat domains, these include roles in vesicular transport and membrane tethering, both of which could speculatively be relevant for ER membrane-based GPI biosynthesis (Striegl *et al*., 2010).

Tb927.10.7130 is a hypothetical protein with no predicated signal peptide or transmembrane domain. The localisation of N and C-terminal tagged Tb927.10.7130 suggests it residues at cortical cytoskeleton (Dean *et al*., 2017) which indicates this protein is unlikely involved in the GPI anchor biosynthesis pathway in the ER.

Tb927.10.2020 is hexokinase which is cytoplasmic and involved in glycolysis and other carbohydrate metabolic processes. Hexokinase is an abundant protein (29^{th} ranking out of 7149 detected proteins) and, like the mitochondrial HSP70, unlikely to be specifically enriched with TbGPI3-3Myc. In summary, the candidate hit most worthy of any further investigation is Tb927.9.14280. However, confidence in this hit is diminished by the lack of "positive control hits" in the form of one or more of the other known GPI GnT complex components (Table 5.1). In particular the data from Kinoshita's group on the mammalian GPI GnT suggest that, by analogy with the PIGA, C and H subcomplex, TbGPI3 ought to be physically associated with TbGPI2 and TbGPI15 (Hong et al, 1999). The reasons for their absence could be due to any or all of the following factors: i, The experimental conditions may have been sub-optimal for

cross-linking, in particular (Fig 5.6A) suggests that less of the TbGPI3-3Myc containing protein complexes were formed in the scaled up cross-linking experiment compared with the small-scale trial experiment (Fig 5.5B). This indicates that the cross-linking reaction in the scaled-up experiment was insufficient and the TbGPI3-3Myc containing complexes that were not stabilised by the crosslinker would be disrupted with stringent 2% TX-100 solubilisation following the cross-linking reaction. ii, The efficacy of cross-linking reactions to cross link the specific partner proteins of TbGPI3 may have been diluted when applied to the whole cell lysate, as opposed to the lower protein content washed membrane fraction. The idea of using whole cell lysate for cross-linking in TbGPI3-3Myc IP experiment was to identify novel soluble co-IP proteins for TbGPI3-3Myc which, on the other hand, may have also introduced abundant soluble proteins (like hexokinase and mitochondrial HSP70) for artefactual cross-linking to TbGPI3-3Myc. iii, Lack of biological replicates for rigorous proteomics identification.

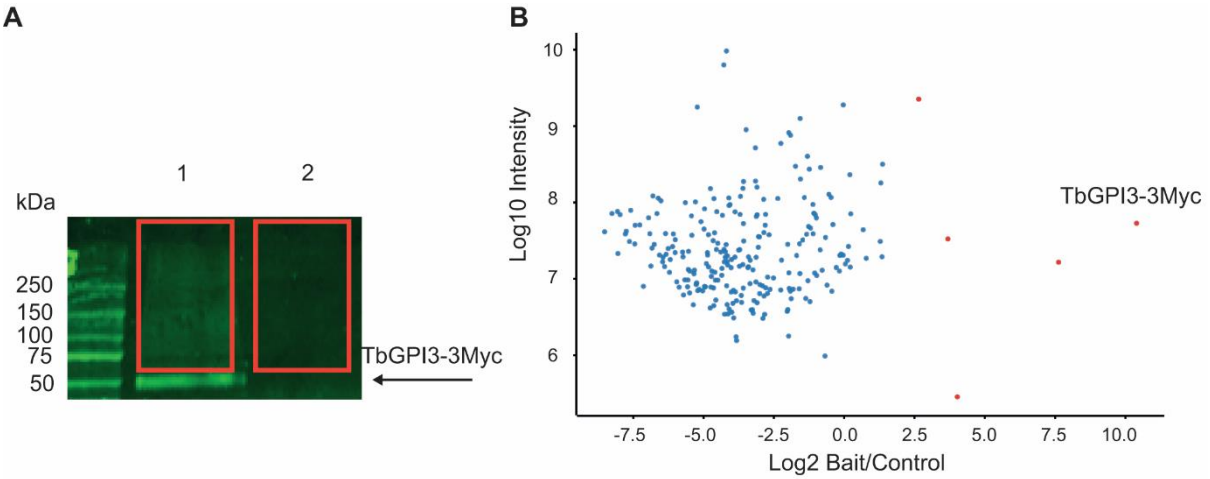

Figure 5.6 Identification of TbGPI3-3Myc co-IP proteins after cross-linking of the hypotonic whole cell lysate.

A: Cross- linking experiments were performed on TbGPI3-3Myc expressing and BSF WT total cell lysates using 1 mM DTSSP. An anti-Myc IP was then performed on 2% TX-100 solubilised proteins. Proteins eluted from the beads were detected by anti-Myc Western blotting. The gel regions (red rectangles) from another gel run in parallel were sent for label free proteomics. B: Plot of proteins identified by LC-MS/MS performed on the regions shown in panel A. X axis: log 2 ratio of protein intensities detected in the TbGPI3-3Myc IP sample to their intensities

detected BSF WT IP samples. Y axis: log 10 value of the sum protein intensities obtained in both TbGPI3-3Myc and BSF WT IP samples.

Table 5.2 List of proteins that co-IP with TbGPI3-3Myc identified by cross linking proteomics

Gene ID	**Log2 bait/control**	**Gene product description**	**Rank in total *T. brucei* cell Proteome**
Tb927.2.1780	10.40	TbGPI3 (N-acetylglucosaminyl-phosphatidylinositol biosynthetic protein, putative)	6475
Tb927.9.14280	7.63	hypothetical protein, conserved	5392
Tb927.10.7130	4.02	hypothetical protein, conserved	3367
Tb927.10.2020	3.69	hexokinase	29
Tb927.6.3800	2.65	heat shock 70 kDa protein, mitochondrial precursor, putative	108

6 Results III - The identification of a *Trypanosoma brucei* GPI anchor inositol deacylase (TbdeAc2) partner protein.

6.1 Identification co-IP proteins of TbdeAc2-3Myc using cross-linking proteomics.

6.1.1 Analysis of *TbdeAc2* Gene Product

The gene product of *Tb927.3.2610*, GPI inositol deacylase 2 (TbdeAc2) was identified as a GPI anchor inositol deacylase in *T. brucei* through BLAST search using rat PGAP1 amino acid sequence as query sequence (Hong, *et al.*, 2006). TbdeAc2 encodes a protein of 815 amino acids (91.4 kDa theoretical MW) with 11 predicted transmembrane domains (Käll, Krogh and Sonnhammer, 2004). Like its yeast homologue Bstp1 (Tanaka *et al.*, 2004), it contains an ~50 amino acid extension at its N-terminus compared to PGAP1. Knockdown of TbdeAc2 by RNAi (Hong, Nagamune, Morita, *et al.*, 2006) showed that that TbdeAc2 removes the inositol-linked acyl chain and that this needs to occur before fatty acid remodelling and attachment of the GPI anchor to VSG in BSF *T. brucei*. These data, and the observation that GPI intermediates from Man_1GlcN-PI onwards are in equilibrium between their inositol acylated and inositol deacylated forms, were consistent with those previously reported using chemical inhibition of inositol deacylation with diisopropylfluorophosphate (DFP) (Güther and Ferguson, 1995). The structure and biosynthesis of GPI anchor intermediates in PCF trypanosomes was described by Field and colleagues (Field, Menon and Cross, 1991)(Field, Menon and Cross, 1991)(Field, Menon and Cross, 1992) and these studies showed that inositol deacylation does not occur in this lifecycle stage. Consistent with this, the TbdeAc2 mRNA transcript levels were found to be 6-fold lower in PCF than in BSF cells (Hong, Nagamune, Morita, *et al.*, 2006).

The fact that TbdeAc2 catalyses several deacylation steps in the BSF GPI anchor biosynthesis pathway, i.e., removing the acyl chain from Man_1GlcN-(acyl)PI, Man_2GlcN-(acyl)PI, Man_3GlcN-(acyl)PI and from glycolipid C (Fig 1.10) suggests that it may physically interact with one or several of the enzymes catalysing adjacent biosynthetic steps. Further, it may require chaperone and/or regulatory proteins for its function. In particular, since TbdeAc2 action is a pre-requisite for fatty acid remodelling, we considered it possible that it might be associated with components of that machinery, such as the unknown component GPI phospholipase A2 (PLA2). For

these reasons we aimed to establish whether TbdeAc2 was present in a protein complex and, if so, identify its partner protein(s).

6.1.2 *In situ* epitope tagging TbdeAc2 and native PAGE of TbdeAc2-3Myc

To investigate whether TbdeAc2 was present in protein complex(es), *in situ* epitope tagging of TbdeAc2 with a 3 × c-Myc epitope tag at its C-terminus was achieved using the same method described in Section 5.2 with primers OL3 and OL4. The success of expressing TbdeAc2-3Myc protein was examined by anti-Myc Western blot. Cell lysates in SDS sample buffer (5×10^6 cell equivalents per lane) were separated by SDS-PAGE and Western blotted with anti-Myc antibody, however, no signal was detected (data not shown). We concluded from this that TbdeAc2 was most likely of very low abundance in BSF trypanosomes, and this is consistent with our recent whole proteome data where TbdeAc2 ranks 5175th in abundance out of 7149 detected proteins (Ji *et al*., 2021) supplementary Table 1). A sequential anti-Myc pull down, similar to that described in Section 4.1.2, was then performed using a 1% TX-100 cell lysates containing 2×10^8 cell-equivalents of the TbdeAc2-3Myc expressing cell line in parallel with identical lysates of the BSF WT cell line. The proteins eluted with SDS sample buffer were then subjected to denaturing SDS-PAGE and analysed by anti-Myc Western blot. TbdeAc2-3Myc was detected in the first pull down at a ~ 90 kDa apparent MW (Fig 6.1A, lane 1), but not in the second pull down (Fig 6.1A, lane 3). As expected, there was no signal detected in BSF WT control samples (Fig 6.1A, lanes 2 and 4). These data indicated that under the conditions used, TbdeAc2-3Myc proteins could be quantitatively immunoprecipitated by a single pull down.

To enable later native-PAGE analysis of TbdeAc2-3Myc IP products, a mild lysis condition using 0.5% digitonin as detergent for pull down and a compatible mild elution method, with c-Myc peptide, were assessed. Anti-Myc Western blot after the pull down showed that TbdeAc2-3Myc could be successfully solubilised by 0.5% digitonin, bound to anti-Myc beads and eluted with either c-Myc containing buffer (Fig. 6B, lane1) or SDS sample buffer (Fig 6B, lane2), albeit in lower yield than using 1% TX-100 lysis and SDS sample buffer elution. The c-Myc peptide eluted samples were then analysed by native-PAGE following by anti-Myc Western blot. As expected, nothing was detected in the sample for the WT control cell lysate (Fig. 6C, lane 1), whereas TbdeAc2-3Myc was clearly detected in a single band at with an appMW of ~ 240 kDa (Fig. 6C, lane 1). This appMW is 2-3 times higher than that of a TbdeAc2 monomer,

indicating it was present as either a homodimer or homotrimer, or a in complex with another protein or proteins.

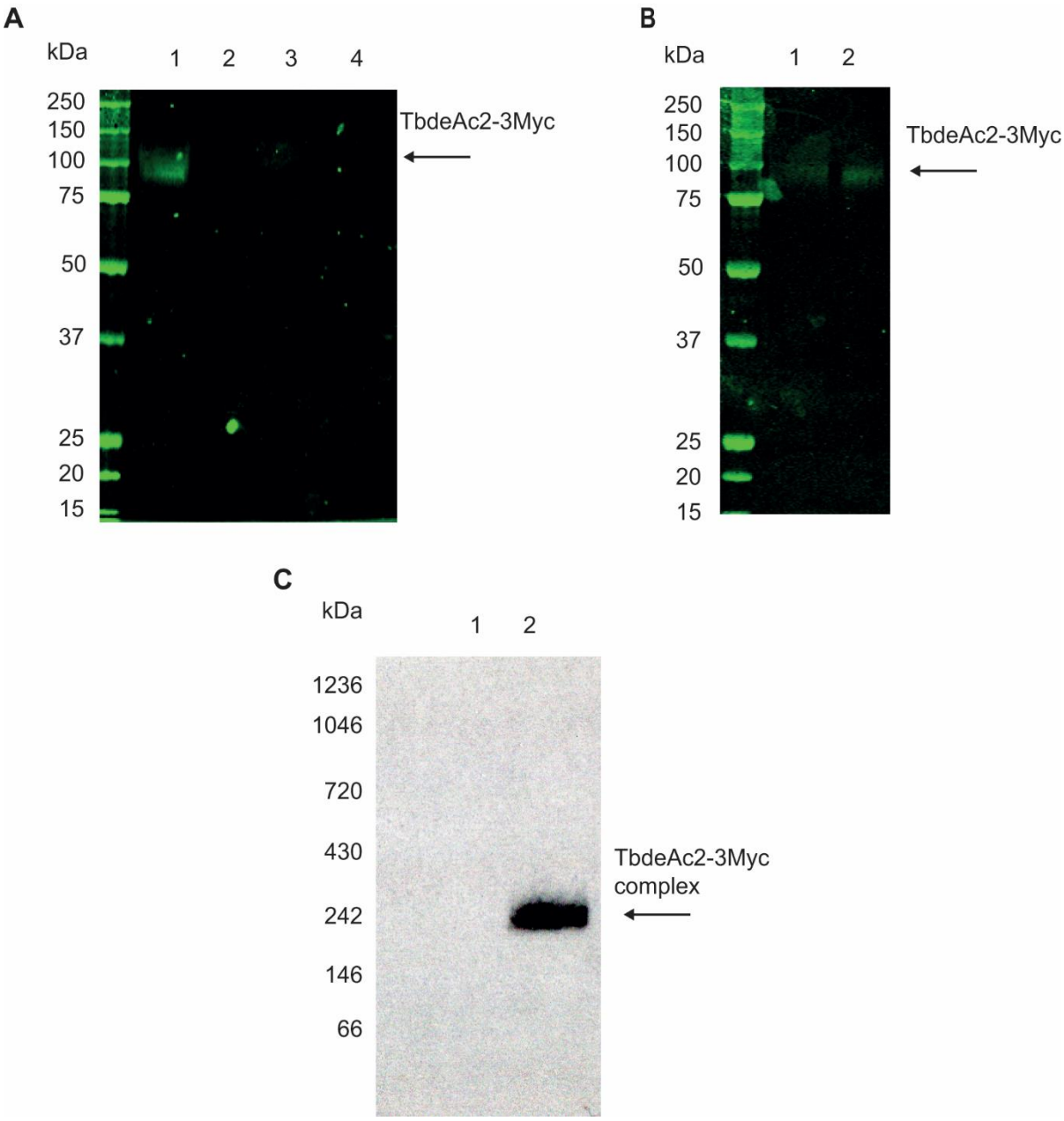

Figure 6.1 Immunoprecipitation and anti-Myc Western Blot of TbdeAc2-3Myc.

A: Anti-Myc Western blot analysis of the IPs from TbdeAc2-3Myc expressing (lane 1 and lane 3) and BSF WT (lane 2 and 4) *T. brucei* 1%TX-100 cell lysates following two sequential IPs with anti-Myc agarose beads. B: TbdeAc2-3Myc expressing cells were lysed with 0.5% digitonin followed by anti-Myc pull down. Proteins were eluted with 0.5% digitonin buffer containing 1 mg/mL c-Myc peptide (lane 1) or SDS sample buffer (lane 2) and subjected to reducing SDS-PAGE and anti-Myc immunoblotting. C: A duplicate IP to that described in (panel B), and a control IP made using equivalent 0.5% digitonin WT cell lysate, were eluted

with c-Myc peptide and subjected to native PAGE and anti-Myc Western blot analysis. Lane 1, sample from WT control lysate; lane 2 from TbdeAc2-3Myc expressing cell lysate.

6.1.3 Attempt to identify co-IP proteins with TbdeAc2-3Myc by label-free proteomics

Having established appropriate pull-down conditions that preserve the protein complex that TbdeAc2-3Myc forms, label free quantitative proteomics were performed with the TbdeAc2-3Myc and BSF WT IPs to further identify the components within the complex. For this experiment, BSF WT and TbdeAc2-3Myc expressing parasites were grown under identical conditions and the same numbers of cells were harvested and lysed in 0.5 % digitonin lysis buffer. Pull down was performed with the cell lysates and proteins were eluted with SDS sample buffer, resolved by SDS-PAGE, gel pieces were cut and digested with trypsin and analysed by LC-MS/MS. The proteins identified from proteomics were plotted by the Log10 value of their protein intensity detected by mass spectrometry (y axis) against the log2 value of the ratio of their intensities in the TbdeAc2-3Myc IP sample versus in the BSF WT IP sample. As shown in (Fig 6.2), the most highly enriched protein was Tb927.11.17650 which is a truncated variant surface glycoprotein sequence of 134 amino acids with an N-terminal signal peptide. However, it was more enriched than the TbdeAc2-3Myc bait protein, so we expect this is an artefact. Although TbdeAc2-3Myc was the second most enriched protein, there were another109 proteins detected with log 2 bait/control values greater than 5, and 465 proteins with log2 bait/control values greater than 2 (https://github.com/mtinti/TbdeAc2attempt1). It seemed very unlikely that we were seeing selective enrichment of proteins with TbdeAc2-3Myc in this experiment, so we investigated alternative approaches. If we were to return to this label-free approach, we would use c-Myc peptide eluates rather than SDS sample buffer eluates to reduce the background.

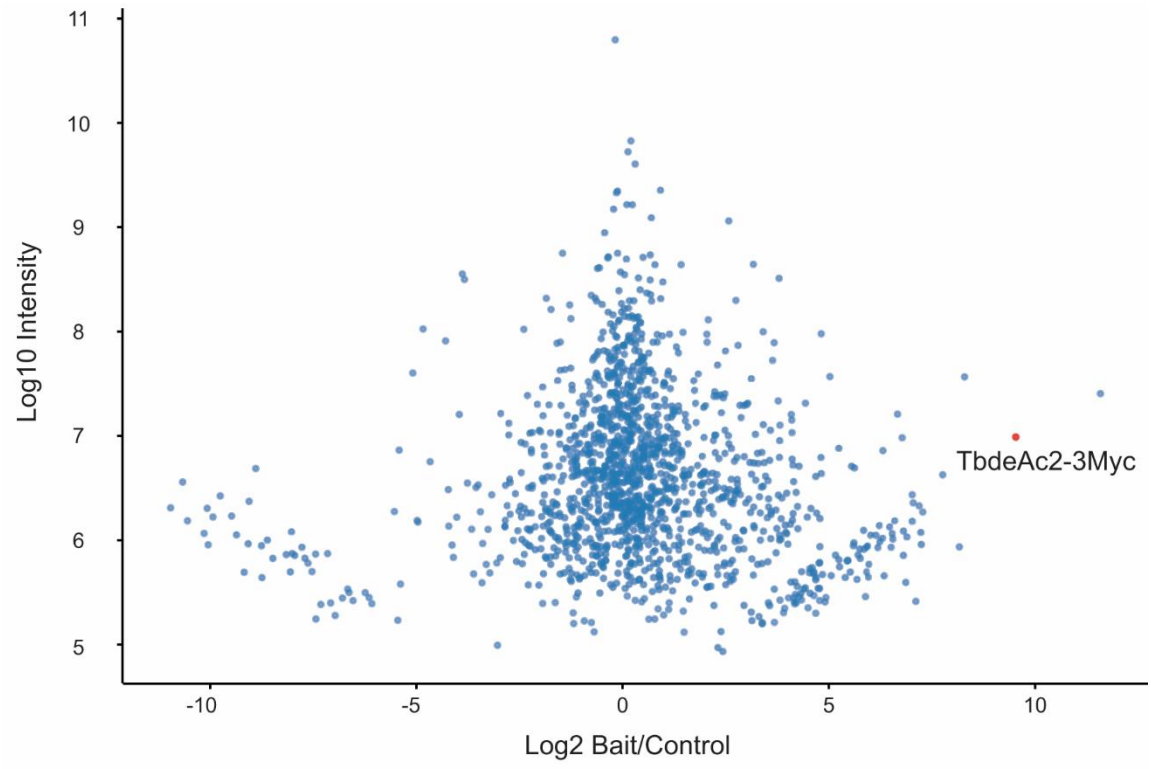

Figure 6.2 Identification of co-IP proteins of TbdeAc2-3Myc by label-free quantitative proteomics.

Plot of TbdeAc2-3Myc versus BSF WT IP results. The plot shows the Log2 value of the intensity of proteins in the bait (TbdeAc2-3Myc) IP relative to the control (BSF WT) IP (x axis) against the Log10 value of protein intensity (y axis). The TbdeAc2-3Myc (bait) protein is annotated and highlighted in red.

6.1.4 Identification of co-IP proteins with TbdeAc2-3Myc through cross-linking proteomics

As TbdeAc2 is a transmembrane protein, and the partner proteins associated with TbdeAc2 may also contain hydrophobic domains, the cross-linking method that was developed for studying transmembrane protein complexes (Section 4.2) was applied to TbdeAc-3Myc. Briefly, BSF *T. brucei* cells expressing TbdeAc2-3Myc were subjected to hypotonic lysis and the membrane fraction from the lysate was collected by centrifugation. Cross-linking reactions were performed on the re-suspended membrane fractions by adding different concentrations of DTSSP. The proteins were then solubilised with 2% TX-100 followed by anti-Myc immunoprecipitation, the proteins eluted from the beads with SDS sample buffer were analysed by either immunoblotting or mass spectrometry.

First, different concentrations of DTSSP were tested to define the optimal concentration for the TbdeAc2-3Myc cross-linking reaction. From the test reactions (0, 1, 2, 4 and 8 mM DTSSP), DTSSP-dependent high-molecular weight smeared bands appeared in a DTSSP concentration-dependent manner (Fig. 6.3A, lanes 1-5). Based on these results, we selected 4 mM DTSSP for the following proteomics experiments, based on similar considerations described in (Section 4.2). Scaled-up the cross-linking experiments using 4 mM DTSSP were performed on washed cell membranes from the TbdeAc2-3Myc expressing cell line and from control WT cells. The cross-linked membrane pellets were solubilised and anti-Myc pull downs were performed as previously described (Section 4.2). The eluted proteins were separated by SDS-PAGE and analysed by either anti-Myc Western blot (Fig 6.3B) or LC-MS/MS. The gel regions that were sent for proteomics are illustrated in the red rectangle regions in (Fig 6.3B).

The proteomics data were plotted as the ratio of the intensity of the protein in the TbdeAc2-3Myc tagged sample divided by the sum intensity of the protein in both samples (x axis) and the log10 value of the protein scores. In this analysis, the proteins that were specifically enriched (i.e., TbdeAc2-3Myc and its partner proteins) should have a ratio that is very close to 1. A cut off-off of protein scores >15 was also applied for this dataset. Under these criteria, there were two protein groups identified with ratio of 1 (Fig 6.3C). The one with a higher protein score was the bait protein TbdeAc2-3Myc, the other protein(s) was/were Tb927.11.11740 and/or Tb927.11.11750 which are too similar in sequence to be discriminated by the LC-MS/MS data. One or both of these proteins were considered to be likely candidate partner protein(s) for TbdeAc2.

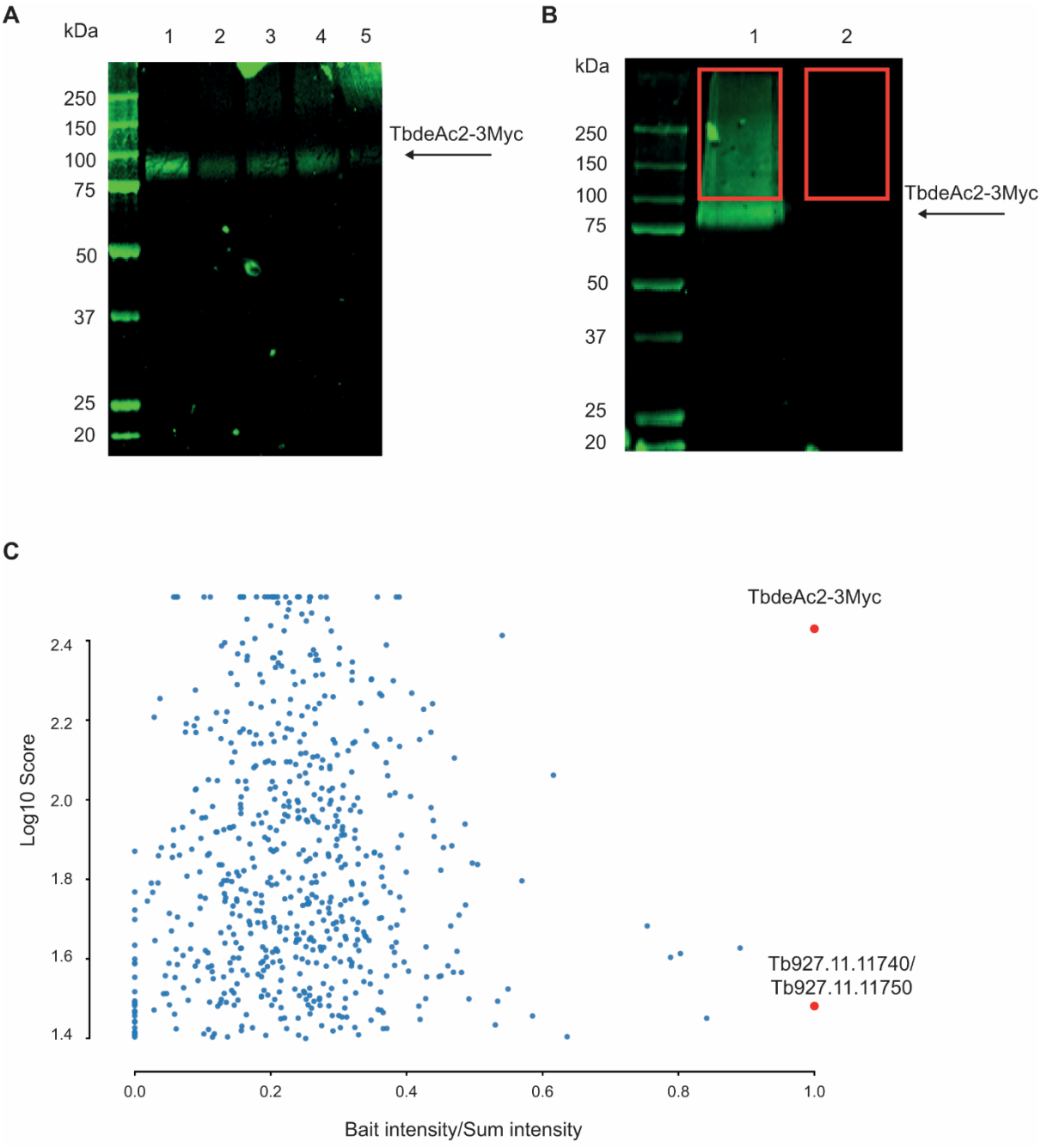

Figure 6.3 Cross-linking proteomics of TbdeAc2.

A, Optimisation of cross-linker concentrations for TbdeAc2-3Myc. TbdeAc2-3Myc expressing cells were lysed and the membrane fraction was collected and cross-linking reactions were performed using 0, 1, 2, 4 and 8 mM DTSSP (lanes 1-5, respectively). Membrane proteins were then solubilised in 2% TX-100 and IPs were performed. Eluted proteins were separated on SDS PAGE and anti-Myc Western blotting was performed on the transferred membrane. B, Cross-linking experiments were performed on TbdeAc2-3Myc expressing (lane 1) and BSF WT (lane 2) washed cell membranes using 4 mM DTSSP. An anti-Myc Western blot was performed to on the proteins that were eluted after IP. Duplicate gel regions to those indicated in red rectangles were sent for proteomics. C, Plot of label-free quantitative proteomics data obtained from the regions shown in panel B. X axis: ratio of protein intensities detected in the TbdeAc2-3Myc IP

sample compared with the sum of their intensities detected in the TbdeAc2-3Myc and BSF WT samples. Y axis: log 10 value of the Andromeda protein score calculated by Maxquant.

6.2 Identification co-IP proteins of TbdeAc2-3Myc by SILAC proteomics

6.2.1 Identification of co-IP proteins of TbdeAc2-3Myc with 0.5% digitonin buffer as lysis condition for pull down

To validate the co-IP proteins of TbdeAc2-3Myc identified by cross-linking proteomics, an orthoganal pull-down experiment was then carried out under very mild solubilisation conditions and using SILAC proteomics methodology. For this experiment (Fig 6.4A), BSF WT and transgenic parasites (expressing TbdeAc2-3Myc) were grown under identical conditions for nine cell divisions, except that the TbdeAc2-3Myc expressing cell line was grown in "heavy medium" containing stable isotope labelled Lys and Arg (R_6K_4), while the BSF WT cells were grown in "light medium" containing unlabelled Lys Arg (R_0K_0). The BSF WT and the transgenic TbdeAc2-3Myc cells were counted and mixed together in a 1:1 ratio and lysed in 0.5% digitonin buffer after harvesting and washing steps. After immunoprecipitation by Myc-Trap™ beads, the proteins eluted with c-Myc peptide were processed to tryptic peptides for LC-MS/MS analysis. In this SILAC experiment, TbdeAc2-3Myc and any proteins that were specifically associated with it can be distinguished from non-specific proteins by the isotope ratios of their tryptic peptides: TbdeAc2-3Myc and its true partner protein peptides will have high heavy/light isotope ratios whereas non-specific proteins will have roughly equal heavy/light isotope ratios. This experiment was performed in biological triplicates to provide statistical power. The raw data from this SILAC experiment was used to search a *T. brucei* predicted protein database using MaxQuant software. Each protein was displayed on a plot of the mean of the minus log10 p value of a t-test (between the TbdeAc2-3Myc and the BSF WT sample) (y axis) and the log2 value of the heavy to light isotope ratios of the same peptides (x axis) (Fig 6.4B). The bait protein TbdeAc2-3Myc protein had the highest heavy/light ratio (14:1), and there was only one other protein group that was significantly enriched (i.e., with a heavy/light ratio > 4). This corresponded to Tb927.11.11740 or/and Tb927.11.11750 with a heavy/light ratio of 6. The reproducibility of the SILAC experiment was also visualised with three different types of pair-plots (histogram, scatter and density plots) using the log2 value of the heavy to light isotope ratios of the protein peptides collected from each replicate (Fig

6.4C). The Pearson correlation coefficient values calculated for each pair were very close to 1 (0.88 for replicates 1 and 2, 0.88 for replicates 1 and 3 and 0.90 for replicates 2 and 3), indicating these three triplicates were highly correlated with each other. The identification of Tb927.11.11740 and/or Tb927.11.11750 as the co-IP protein(s) of TbdeAc2 through both cross-linking and SILAC proteomics analyses strongly supported the conclusion that TbdeAc2-3Myc interacts with Tb927.11.11740 and/or Tb927.11.11750. To examine whether both or only one of Tb927.11.11740 and Tb927.11.11750 is/are true partner protein(s) of TbdeAc2-3Myc, the two proteins were studied individually in later pull-down experiments.

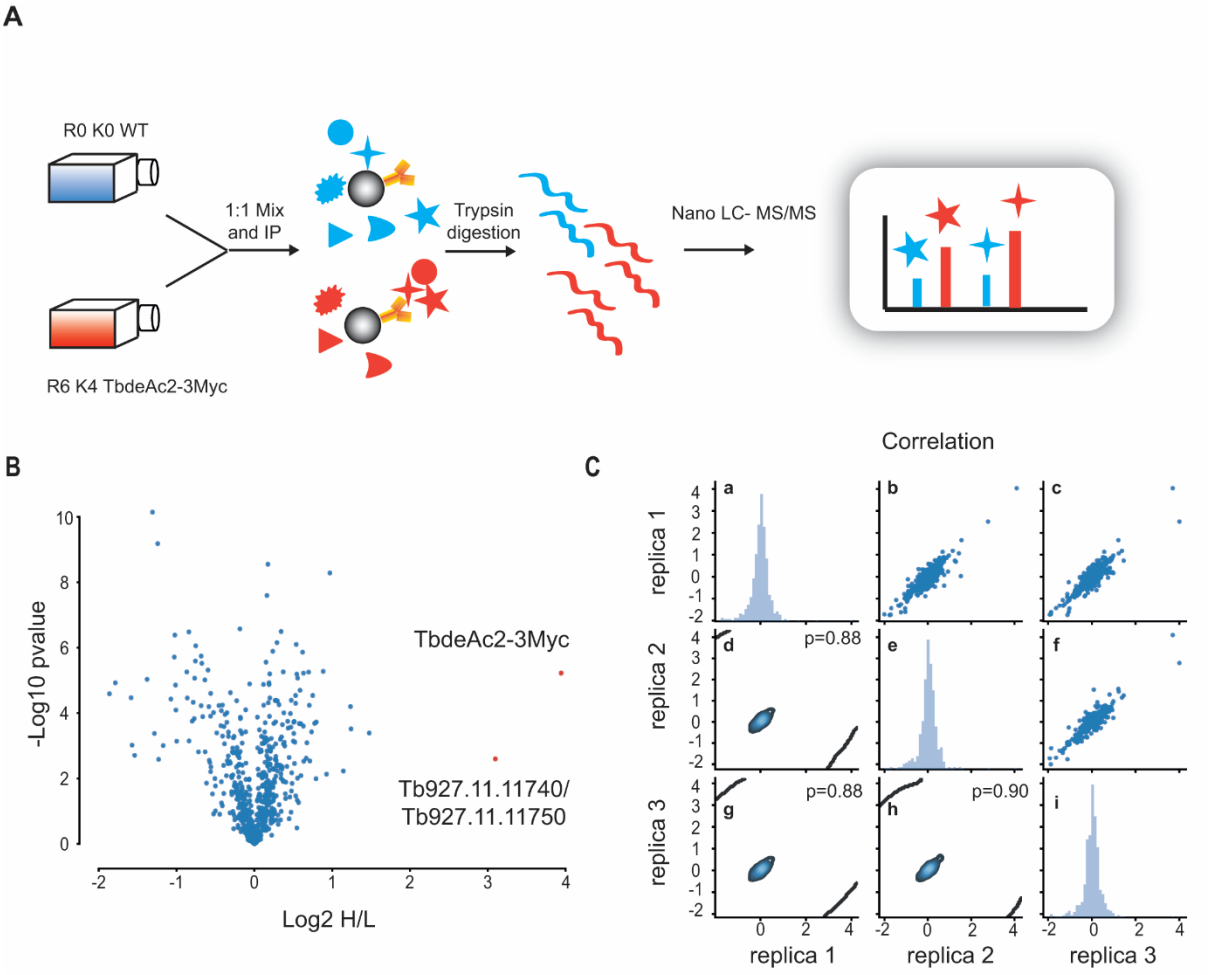

Figure 6.4 Summary of SILAC proteomics experiment.

A: Scheme of SILAC proteomics to identify proteins that specifically co-IP with of TbdeAc2-3Myc. B: Volcano Plot of the combined SILAC proteomics data from 3 biological replicates. X axis: log 2 value of heavy to light isotope ratios of proteins peptides. Y axis: mean of minus log10 value of p values from a t-test. C: Reproducibility analysis of three replicates from SILAC proteomics data. A pairwise comparison of the log2 value of the heavy to light isotope ratios of the protein peptides of three replicates visualizing with histogram counts (a, e, i) scatter

plots (b, c, f) and density plots (d, g, h). The Pearson correlation coefficient values (abbreviated as p) of each pair are also reported on the density plots.

6.2.2 Attempt to find additional co-IP proteins of TbdeAc2-3Myc using 0.3% digitonin containing lysis buffer for pull down

When performing pull down experiments coupled with SILAC proteomics identification methods searching for the co-IP proteins of TbdeAc2, another milder lysis condition which utilises 0.3% instead of 0.5% digitonin containing buffer was tested for protein solubilisation for pull down. All the other procedures were identical as described in section 6.2.1 and this experiment was also performed in biological triplicates. The proteins that were identified were plotted by mean of minus log10 p value of a t-test (between the TbdeAc2-3Myc and the BSF WT sample) (y axis) and the log2 value of the heavy to light isotope ratios of the same peptides (x axis) (Fig 6.5). TbdeAc2-3Myc protein had the highest heavy/light ratio which was 16:1, while the other protein whose heavy/light ratio was higher that 4 (with ratio 4.6:1) was Tb927.4.4580 which was not the same co-IP protein identified in previous two proteomics methods. Tb927.4.4580 is a hypothetical protein ranking the 2711th out of 7149 identified proteins based on the BSF whole proteins proteomics dataset (Ji, Tinti and Ferguson, 2021) which indicated that it was a relatively abundant protein that may have been non-specifically bound to TbdeAc2-3Myc. According to TritrypDB (Aslett *et al.*, 2010), Tb927.4.4580 is a cytoplasmic protein that does not contain signal peptide or transmembrane domain. There are no known motifs or domains annotated in this protein either. The features of this protein suggests that it is not involved in the GPI anchor biosynthesis pathway, thus we did not perform further studies on it. In this experiment Tb927.11.11740/50 did not appear at all in the pull-down dataset, perhaps suggesting that 0.3% digitonin failed to liberate it from the ER membrane.

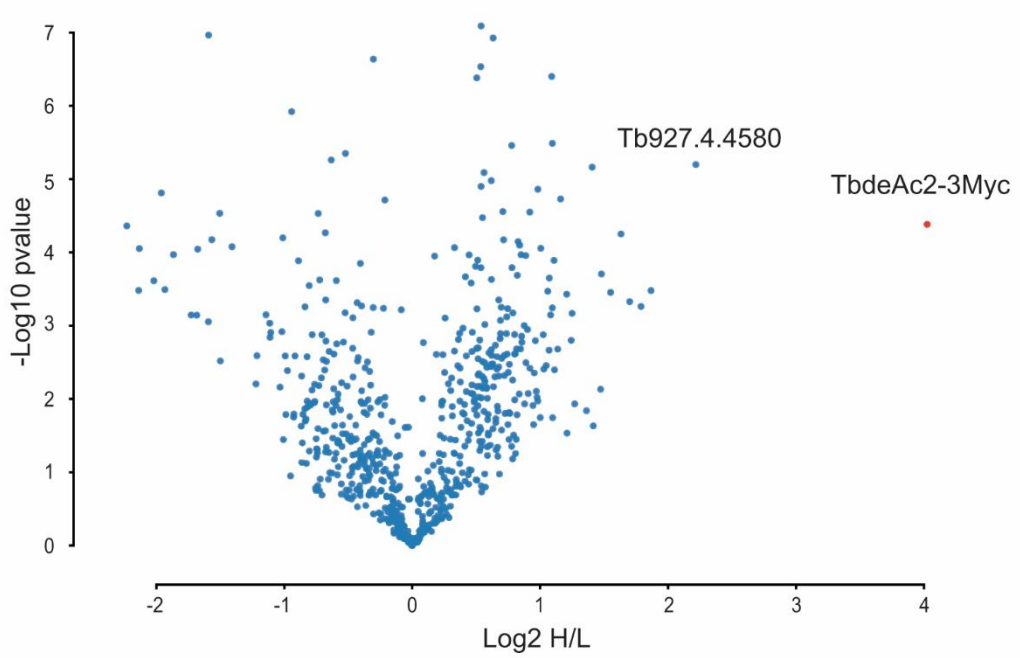

Figure 6.5 Volcano Plot of the combined SILAC proteomics data from 3 biological replicates.

X axis: log 2 value of heavy to light isotope ratios of proteins peptides. Y axis: mean of minus log10 value of p values from a t-test.

6.3 Identification of Tb927.11.11750 as the sole co-IP protein of TbdeAc2 through reverse pull down

6.3.1 Analysis of Tb927.11.11740 and Tb927.11.11750

The co-IP protein(s) of TbdeAc2 found by two different proteomics approaches is/are a hypothetical protein(s) encoded by one or both of two closely-related tandemly arrayed genes. These two genes are *Tb927.11.11740* and *Tb927.11.11750* which share 94.5% identity in nucleic acid sequence, further, 248 bp of 5' untranslated region (UTR) sequence upstream of the start codons of both genes are identical. Analysis of the predicted amino acid sequences of these two proteins show they share significant similarities but some differences: The alignment of amino acid sequences of Tb927.11.11740 and Tb927.11.11750 (Fig 6.6A) shows they share 92 % identity in amino acid sequence but with 52 amino acid differences at their C-termini. This high degree of sequence identity is why the proteomics methods could not distinguish these two proteins based on tryptic peptide mass spectrometry as no unique C-terminal peptides of these two proteins were identified. Both proteins contain a 24 amino-acid (aa) predicted signal peptide (SP) at their N-termini (Kall *et al*., 2004) (Käll, Krogh and

Sonnhammer, 2004) and one transmembrane domain near their C-termini (Käll, Krogh and Sonnhammer, 2004). In addition, both proteins have a putative catalytic lipase motif GXSXG (Derewenda and Sharp, 1993) and a potential N-glycosylation site (Gupta and Brunak, 2002). However, only protein Tb927.11.11750 has a possible ER retention signal (KXKXX)(Kruzel *et al*., 2017) at its C-terminus, indicating that Tb927.11.11740 and Tb927.11.11750 may have different subcellular localisations. There is also a phosphoglycerate mutase-like Histidine phosphatase superfamily domain (SSF53254) (Wilson *et al.*, 2007) found in both proteins. The orthologs of both two genes have been only identified in the kinetoplastidea class (*Trypanosoma theileri*, *Trypanosoma Cruzi* ect) indicating they likely emerged at the same time point in the evolutionary progress. To investigate the associations of these two proteins with TbdeAc2 individually, epitope tagging of Tb927.11.11740 and Tb927.11.11750 was achieved through *in situ* tagging and internal tagging, respectively.

A

```
Tb927.11.11740      MMRPLSLVWWAIFVLLAATGANCSYVLELVQVVHRGGITPPPVGTPDREKLCSPDSGSSC      60
Tb927.11.11750      MMRPLSLVWWAIFVLLAATGANCSYVLELVQVVHRGGITPPPVGTPDREKLCSPDSGSSC      60
                    ************************************************************

Tb927.11.11740      AAIANHGVQQLIDMGAYIEKLYKNDDETEGAKAWLGSTYDSAAVYTHSFADPALVQAATA      120
Tb927.11.11750      AAIANHGVQQLIDMGAYIEKLYKNDDETEGAKAWLGSTYDSAAVYTHSFADPALVQAATA      120
                    ************************************************************

Tb927.11.11740      LLKGIYAEEQGNITPAVISAPPADDTLLNVNALPSFVLGNEAKAQHFNKTMEEAVDEQFP      180
Tb927.11.11750      LLKGIYAEEQGNITPAVISAPPADDTLLNVNALPSFVLGNEAKAQHFNKTMEEAVDEQFP      180
                    ************************************************************

Tb927.11.11740      DSSVIGTMGREVGLSEACSSADNRVWCCHRLQQLATMYSAMKDGGSGAPTVMENVGRLDA      240
Tb927.11.11750      DSSVIGTMGREVGLSEACSSADNRVWCCHRLQQLATMYSAMKDGGSGAPTVMENVGRLDA      240
                    ************************************************************

Tb927.11.11740      VATVRSHTLYGYSTEDELAKARGSLGQPLAQELLLGMRRKMLQKDDVNYNTHKVMQYVHN      300
Tb927.11.11750      VATVRSHTLYGYSTEDELAKARGSLGQPLAQELLLGMRRKMLQKDDVNYNTHKVMQYVHN      300
                    ************************************************************

Tb927.11.11740      TPIRTTVGYQPTDLVPVAETFLIDLLRDDTTEVYFVRLRHAVVKTEPAAMPSAADFPFRC      360
Tb927.11.11750      TPIRTTVGYQPTDLVPVAETFLIDLLRDDTTEVYFVRLRHAVVKTEPAAMPSAADFPFRC      360
                    ************************************************************

Tb927.11.11740      VNAAGDSSGVKSKDGFICPFDDFVRFVDSSKGTGAGGAICHLDTSTADNLQCSVVGPPTS      420
Tb927.11.11750      VNAAGDSSGVKSKDGFICPFDDFVRFVDSSKGTGAGGAICHLDTSTADNLQCSVVGPPTS      420
                    ************************************************************
                                                448
Tb927.11.11740      PQCERYRRLCPRLACPDGHIYDVVSGMCSLRFVKVDAMRSDGAIGLIVSMLFIGVILSIG      480
Tb927.11.11750      PQCERYRRLCPRLACPDGHIYDVVSGMCTSLYADEAILTGGVVASLCIALLFSGVMLGII      480
                    ****************************:  :..   : .. . .* :::** **:*.*

Tb927.11.11740      ILHWCPLLCLGVSPKDPTVR        500
Tb927.11.11750      MVEMYPVMFKQKMAV-----        495
                    ::.  *::
```

B

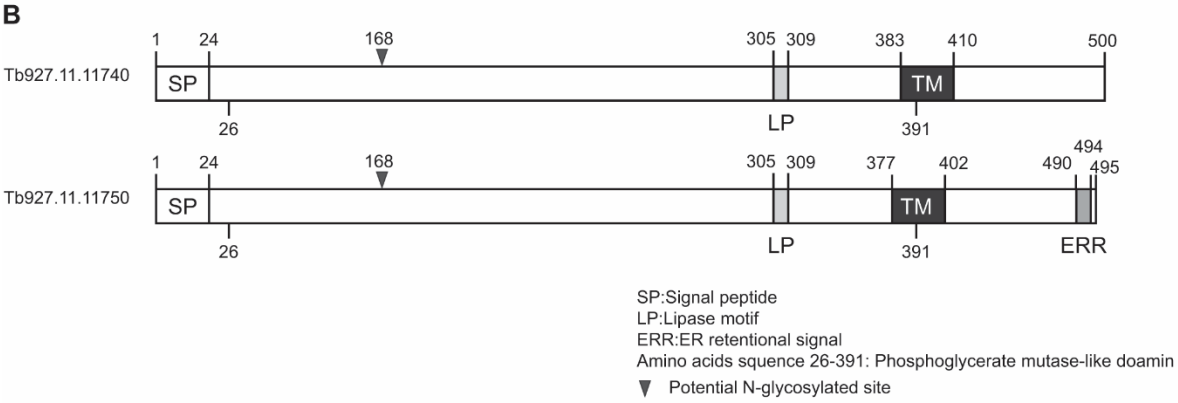

Figure 6.6 Summary of protein features of Tb927.11.11740 and Tb927.11.11750.

A: Alignment of amino acid sequences of Tb927.11.11740 and Tb927.11.11750 generated by T coffee. * = identical residues, : = conserved residues. Protein sequences were retrieved from TritrypDB dataset (Aslett *et al.*, 2010). Amino acid residues 1-448 of these two proteins are identical. B: Comparison of Tb927.11.11740 and Tb927.11.11750 proteins. Different domains and motifs are indicated. Amino acid numbers are above the models.

6.3.2 Tb927.11.11740 does not form a protein complex with TbdeAc2

To independently investigate whether Tb927.11.11740 forms complex with TbdeAc2, *in situ* tagging of Tb927.11.11740 at its C-terminus with triple human influenza hemagglutinin (HA) tag was achieved using same strategy introduced in Section 5.2 by changing the template plasmid from pMOTag43M to pMOTag4H (Oberholzer *et al.*,

2006) with primers OL5 and OL6. Although *Tb927.11.11740* and *Tb927.11.11750* share very high identity at the nucleic acid sequence level, the unique 3'end coding sequence and 3'UTR sequence of *Tb927.11.11740* makes it possible to *in situ* tag *Tb927.11.11740* without interfering with *Tb927.11.11750*. The *in situ* tagging construct was amplified from pMOTag4H and transfected into BSF WT cell line, hygromycin was added for selection. Recovered cells from the transfection were cloned by limit dilution and three clones picked to be checked by immunoblotting. Aliquots of cell lysates prepared with 1 × SDS sample buffer (equivalent to 5×10^6 cells) of the three clones were subjected to SDS-PAGE and anti-HA Western blot. However, no signal was detected in this experiment (data not shown). This was expected as Tb927.11.11740 is a low abundance protein with a ranking 6882^{th} out of 7149 detected proteins from our quantitative proteomics dataset (Ji, Tinti and Ferguson, 2021). After this test experiment, 5×10^7 Tb927.11.11740-3HA expressing cells of these three clones were then lysed with 1% TX-100 and immunoprecipitation was carried out with anti-HA magnetic beads. After washing, the proteins were eluted with 1 × SDS sample buffer and analysed by denaturing SDS-PAGE followed by anti-HA Western blot. Tb927.11.11740-3HA was detected in all three clones at an apparent molecular weight of ~64 kDa (Fig 6.7A lane 1-3), which was higher than its theoretical molecular weight 54 kDa; the possible reasons for detection of higher molecular weight bands will be explained later. Subsequently, mild lysis and elution conditions for pulling down Tb927.11.11740-3HA were examined. From the anti-HA Western blot of different IP samples separated on a denaturing gel (Fig 6.7B), 0.5% digitonin lysis buffer (lane 2) could solubilise Tb927.11.11740-3HA protein, albeit with less efficiency compared to 1% TX-100 (lane 1), using 1 × SDS sample buffer as the elution condition. Proteins bound to the beads could also be recovered by HA peptide elution buffer (lane 3) under the 0.5% digitonin lysis condition. However, more than half of the HA-tagged protein was left on the beads, as revealed by subsequently boiling the same beads in 1 × SDS sample buffer (lane 4). No HA-tagged proteins or anti-HA cross-reacting proteins were detected in the BSF WT samples, demonstrating the specificity of the pull down and detection system (lane 5). To check whether Tb927.11.11740 was present in any protein complexes, 6×10^8 BSF WT and Tb927.11.11740-3HA expressing cells were lysed with 0.5% digitonin containing buffer and immunoprecipitation was performed on the supernatants after centrifugation. Proteins that bound to the beads were eluted with 1 mg/mL HA peptide containing buffer. In this experiment, duplicated samples of each cell line were prepared, one duplicate was subjected to denaturing SDS-PAGE and the

other was analysed by native-PAGE followed by anti-HA Western blot. The blot of the denaturing gel (Fig 6.7C, lane 1) showed a stronger detection of Tb927.11.11740-3HA compared to the 5×10^7 cell lysate with no detection in BSF WT sample. The duplicate samples on the native-PAGE (Fig 6.7D) showed that Tb927.11.11740-3HA was detected in two dominant bands. The stronger band that was at 66 kDa was similar to the apparent molecular weight of the protein monomer detected in the denaturing gel while the other band migrated between 146 to 242 kDa, suggesting that at least part of Tb927.11.11740-3HA was present in a protein complex.

To ascertain what this protein complex was composed of, and whether TbdeAc2 was within the complex, triplicate IP samples of Tb927.11.11740-3HA and BSF WT were applied to native-PAGE and the gel regions that contained the Tb927.11.11740-3HA protein complexes and its counterparts in the control sample were sent for quantitative label free proteomics identification. The proteins identified from mass spectrometry were plotted by log10 value of their protein peptide intensities (y axis) and the log2 value of the peptide intensity detected in the Tb927.11.11740-3HA sample (bait) versus in the BSF WT (control) sample (Fig 6.8). As expected, the bait protein Tb927.11.11740-3HA was the most highly-enriched protein, however, TbdeAc2 (Tb927.3.2610) was not detected in this experiment indicating that Tb927.11.11740-3HA does not form a protein complex with TbdeAc2. Besides the bait protein, there were nine proteins identified with a log2 bait/control ratio greater than 2 as listed in (Table 6.1). Of these proteins, most of them were very high abundance proteins based on the protein ranking from the whole BSF WT proteomics dataset (Ji *et al*., 2021). These proteins were most likely non-specific proteins co-immunoprecipitated with Tb927.11.11740. Of these: Tb927.10.13800 has a protein ranking 1400^{th} out of 7149^{th} identified proteins (Ji *et al*., 2021) and is therefore a relatively high abundance protein. It has no predicated signal peptide or transmembrane domains so we consider this protein to be unlikely to be a specific partner protein of Tb927.11.11740. Tb927.3.4080 is a putative pyruvate transporter whose protein ranking in the BSF total proteomics data is close to that of the bait protein. It is a 583 amino acid protein containing 14 predicted transmembrane domains and a MFS general substrate transporter domain (SSF103473) (Wilson *et al.*, 2007). This superfamily of proteins are membrane transport proteins that facilitate movement of small solutes across cell membranes in response to chemiosmotic gradients. It is possible that Tb927.3.4080 is a true partner protein for Tb927.11.11740 but, if so, it might suggest that they reside in the plasma

membrane. Their association would need to be verified with more biological replicates before performing function studies.

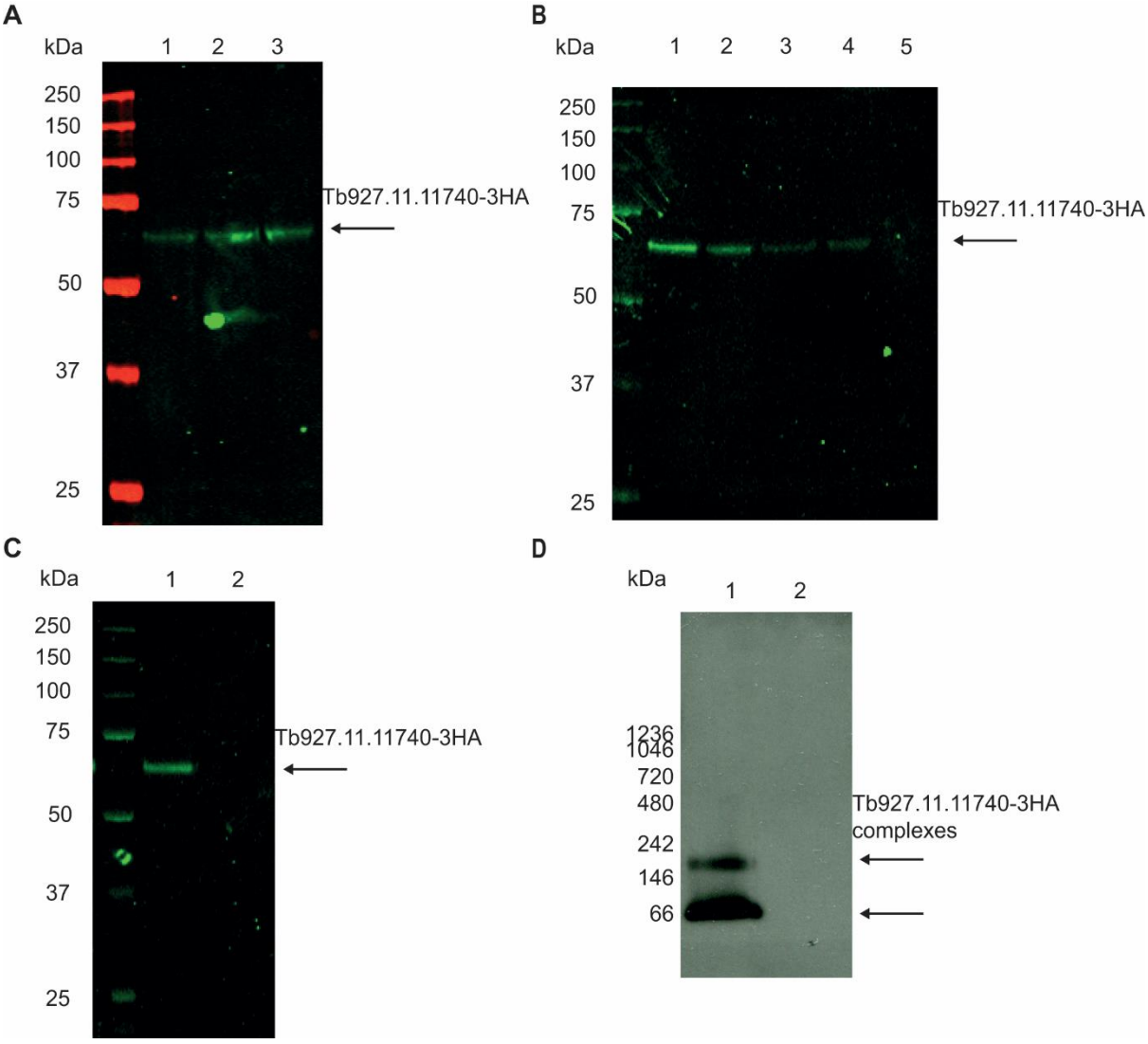

Figure 6.7 Investigation of Tb927.11.11740 binding partners.

A: Anti-HA Western blot of Tb927.11.11740-3HA after pull-down. Three clones of BSF Tb927.11.11740-3HA expressing cells (5×10^7 cells) were lysed in 0.5% digitonin lysis buffer and IP was performed on the supernatant after centrifugation using anti-HA magnetic beads. The beads were washed three times with washing buffer and proteins bound to the beads were eluted with SDS sample buffer and analysed by denaturing gel followed by anti-HA Western blot. B: Test of mild lysis and elution conditions. Tb927.11.11740-3HA was enriched as in panel A except using different lysis and elution conditions. Lane 1: 1% TX-100 lysis buffer and 1 × SDS sample buffer for elution; lane 2: 0.5% digitonin lysis buffer and 1 × SDS sample buffer for elution; lane 3: 0.5% digitonin lysis buffer and 1 mg/mL HA peptide buffer for elution. Lane 4: same beads as lane 3 and sequentially eluted with 1 × SDS sample buffer. Lane 5: BSF WT cell line 1% TX-100 lysate and eluted with 1 × SDS sample buffer. C and D: Tb927.11.11740-3HA and BSF WT cells were immunoprecipitated as in panel B with 0.5%

digitonin as lysis detergent and 1 mg/mL HA peptide solution as elution buffer. Eluted proteins were applied to either denaturing gel (C) or nativePAGE (D) transferred to nitrocellulose (C) or PVDF (D) membranes and detected by anti-HA antibody.

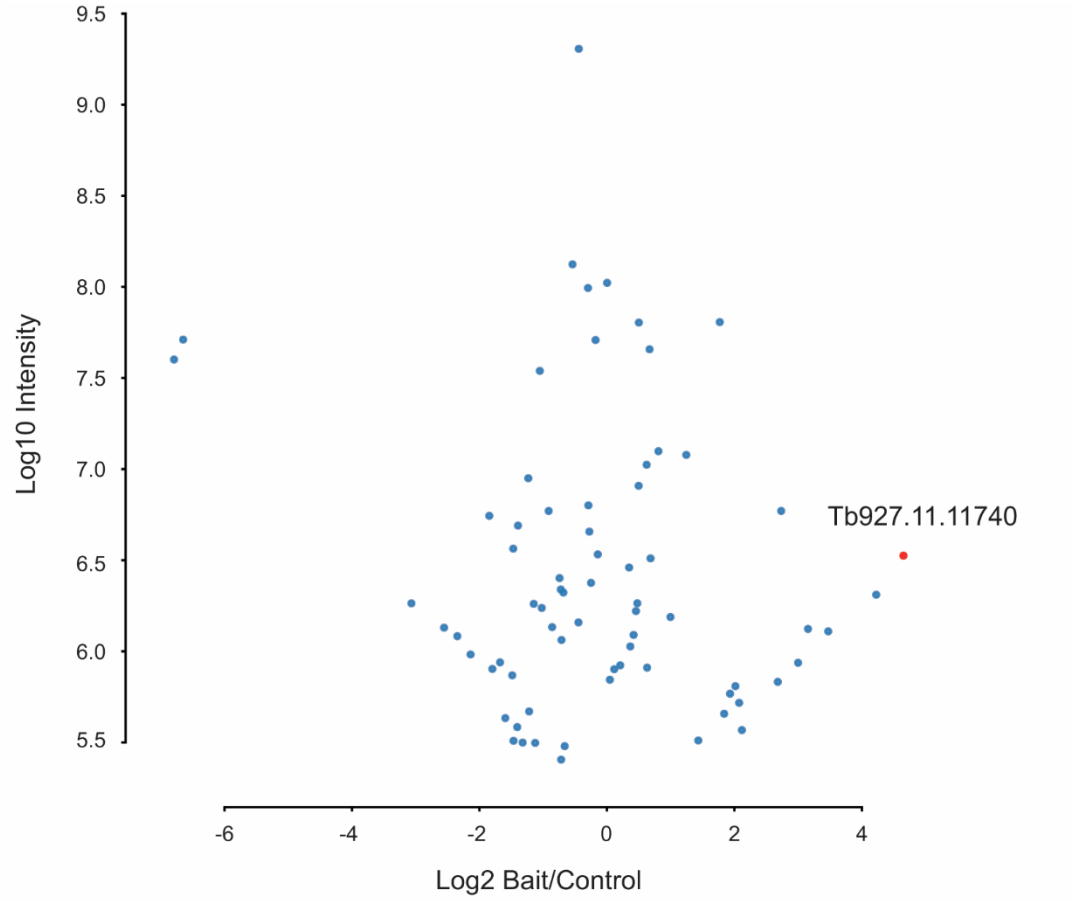

Figure 6.8 Identification of co-IP proteins of Tb927.11.11740 through quantitative label free proteomics.

Plot of proteins identified by LC-MS/MS that were carried out on the high-MW complexes regions of the NativePAGE gels after immunoprecipitating equivalent numbers of Tb927.11.11740-3HA and BSF WT cells. Proteins identified by mass spectrometry were plotted by log 10 intensity of the peptides (y axis) against log2 fold change between the peptides detected in the Tb927.11.11740-3HA bait versus BSF WT control sample.

Table 6.1 List of proteins that co-IP with Tb927.11.11740-3HA identified by label free proteomics.

Gene ID	**Log2 bait/control**	**Gene product description**	**Rank in total *T. brucei* cell Proteome**
Tb927.11.11740	4.651396	membrane-bound acid phosphatase, putative	6882
Tb927.11.5520	4.224686	triosephosphate isomerase	137
Tb927.1.2410	3.472147	beta tubulin, pseudogene	11
Tb927.10.13800	3.154322	hypothetical protein, conserved	1400

Tb927.3.4080	2.998646	Pyruvate transporter, putative	5922
Tb927.9.10770	2.733352	polyadenylate-binding protein 2	133
Tb927.4.2740	2.680354	p25-alpha, putative	294
Tb927.3.3450	2.117101	ADP-ribosylation factor-like protein 3, putative	146
Tb927.3.1380	2.074121	ATP synthase subunit beta, mitochondrial	315
Tb927.11.7380	2.014687	glycerol-3-phosphate dehydrogenase (FAD-dependent), mitochondrial	197

6.3.3 Tb927.11.11740-3HA is N-glycosylated

Previous anti-HA Western blot of Tb927.11.11740-3HA IP products separated on denaturing gel showed that the apparent molecular weight (appMW) of Tb927.11.11740-3HA (~64 kDa) was higher than its theoretical MW (the sum MW of Tb927.11.11740 and triple HA tag is ~ 57 kDa). The prediction of N-glycosylation of Tb927.11.11740 indicated this protein was N-glycosylated (Fig 6.6B) which might be the reason for the detection of extra MW for the protein. To test this hypothesis, digestion was performed using Peptide -N-Glycosidase F (PNGaseF). Briefly, IPs were performed on Tb927.11.11740-3HA and BSF WT 1% TX-100 cell lysates, as described, and after washing steps the proteins bound to the beads were denatured before setting up the digestion with or without PNGaseF, followed by SDS-PAGE and Western blotting with anti-HA antibodies. The blot showed that after PNGaseF treatment, the band corresponding to Tb927.11.11740-3HA shifted below its original detected position (Fig 6.9 A lane1: non treated, lane 2: treated) with no detection in the BSF WT cell controls (Fig 6.9A lane 3: non treated, lane 4: treated). The apparent MW of the detected bands of Tb927.11.11740-3HA were estimated based on the positions and MW of protein markers bands. The correlation of positions of protein marker bands and their molecular weights were plotted by log10 value of the MW (y axis) versus the Rf value of corresponding bands (x axis). In this case, the values of the 100, 75 and 50 kDa bands that were closer to the Tb927.11.11740-3HA bands were chosen to get better correlations. By using this method, the Rf values of Tb927.11.11740-3HA in the PNGase F treated and non-treated samples (0.361 and 0.378) resulted in appMW estimates of ~65 kDa and ~ 61 kDa, respectively. The apparent MW of Tb927.11.11740-3HA in the PNGase F treatment sample (61 kDa) is close to the theoretical MW of 57 kDa indicating that the detection of higher MW band in the non-treated sample was due to the N-glycosylation of Tb927.11.11740.

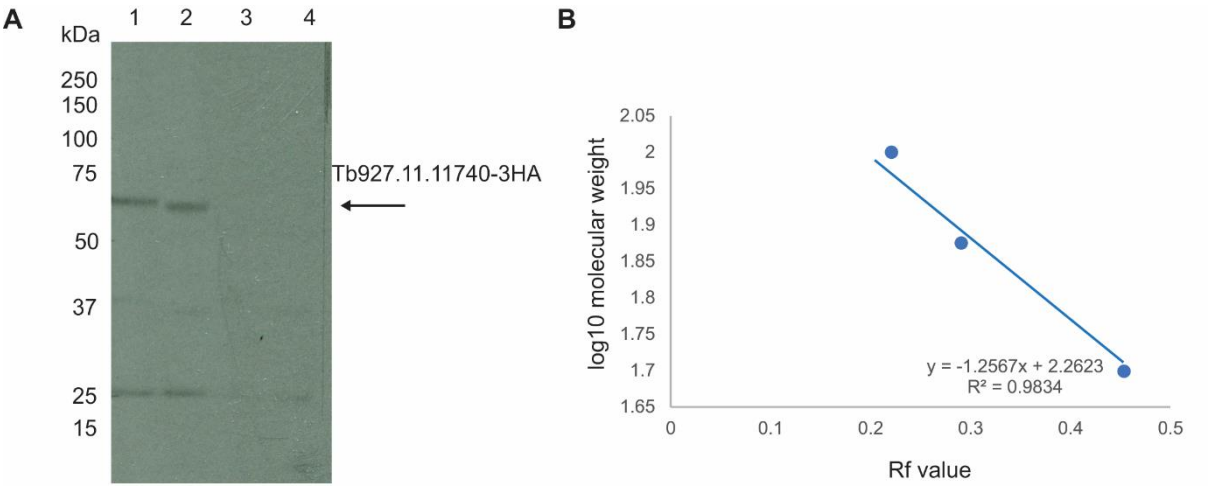

Figure 6.9 Detection of N-glycosylation of Tb927.11.11740-3HA.

A: Pull down experiments were performed with 5×10^7 Tb927.11.11740-3HA and BSF WT 1% TX-100 cell lysates. The proteins that bound to beads were denatured in 1 × Glycoprotein Denaturing Buffer (NEB) at 50 °C for 10 min. The PNGaseF digestions were then performed in 1 × GlycoBuffer(NEB) in the presence of 10% NP-40. The samples were subjected to denaturing SDS-PAGE followed by anti-HA Western blot. (Lane 1 and 2, PNGaseF non-treated and treated Tb927.11.11740-3HA samples; lane 3 and 4, PNGaseF non-treated and treated BSF WT samples). B: Plot of correlations of protein migration positions and molecular weights. X axis: Rf values of proteins (Distance travelled protein divided by distance travelled by the running buffer front). Y axis: log10 value of MWs of proteins. The equation and the R squared value of the trend line were also displayed on the plot.

6.3.4 Attempt to pull down Tb927.11.11750 with specific anti-peptide antibodies

Reverse pull down of Tb927.11.11740-3HA showed that it was not the partner protein of TbdeAc2, leaving the hypothesis that Tb927.11.11750 is the co-IP protein of TbdeAc2. To verify the protein associations between TbdeAc2 and Tb927.11.11750, reverse pull down of Tb927.11.11750 is required. However, the sequence analysis of Tb927.11.11750 shows that it contains a 24 aa signal peptide at its N-terminus and a putative ER retention signal (K(X)KXX) at its C-terminus, making it unwise to tag this protein at its C- or N- terminus. Therefore, we attempted to raise anti-peptide antibodies against unique peptide sequences of Tb927.11.11750. These were raised by two different manufacturers (Davids Biotechnologie and Moravian Biotech) using two different peptide sequences SLYADEAILTGGVVASLC and CMYPVMFKQKMAV. The detection and specificity of these two antibodies were first tested by Western blot against *T. brucei* BSF cell lysates. For the anti-peptide antibody generated by Davids Biotechnologie, 5×10^6 (Fig 6.10 A and B, lanes 1 and 3) and 1×10^7 (Fig 6.10A and B,

lanes 2 and 4) BSF WT cell lysates were prepared and separated on a denaturing gel followed by the probing for Tb927.11.11750 by Western blot using the affinity-purified anti-peptide antibody at 0.5 (Fig 6.10B lanes 1 and 2) and 1 µg/mL (Fig 6.10B lanes 1 and 2). Under these conditions, there was one band detected at ~50 kDa which was close to the theoretical MW of Tb927.11.11750 without its signal peptide. However, very high abundance proteins of *T. brucei* like tubulin and VSGs are also close to the detected bands, thus these detections could be non-specific. Similar detections were also observed when using the anti-peptide antibody raised by Moravian Biotech. In this experiment, 2.5×10^6 cell lysates of BSF WT (Fig 6.10D lanes 1 and 3) and Tb927.11.11750 overexpressing cell line (Fig 6.10D lanes 2 and 4) were tested. The tested overexpression cell line were generated by introducing an ectopic copy of Tb927.11.11750 into BSF WT which was constructed in plew100 original version plasmid (Wirtz *et al.*, 1999). From the Western blot detection using 1 µg/mL (Fig 6.10D lanes 1 and 2) and 2 µg/mL (Fig 6.10D lanes 3 and 4), one band at ~50 kDa was detected in all samples with no significant differences of intensities among them. To check whether this detected band represents Tb927.11.11750 or not, these two different antibodies were coupled to protein A agarose beads to perform pull down experiments. Briefly, lysates equivalent to 1×10^8 cell lysates per sample were prepared with 2%TX-100 lysis buffer, 2 µg of each antibody and protein A agarose beads were added for 1 h incubation, and control samples without antibodies were carried out in parallel. After washing steps, the proteins bound to the beads were eluted and separated on denaturing gel, the gel regions that were expected to contain the Tb927.11.11750 were sent for label free quantitative proteomics identification. The proteins identified by LC-MS/MS were plotted by log10 value of their protein peptide intensities (y axis) and log2 value of the peptide intensities of bait versus control. Unfortunately, Tb927.11.11750 was not detected in either of these pull-down experiments, indicating neither of these two anti-peptide antibodies were able to immunoprecipitate Tb927.11.11750 from *T. brucei* cell lysate. The only highly enriched protein was the IgG protein of rabbit, which suggested that antibodies were bound to protein A agarose beads and that, therefore, the failure in pulling down Tb927.11.11750 was not due to inefficiency of the antibody coupling but poor affinity of the antibodies for their target protein.

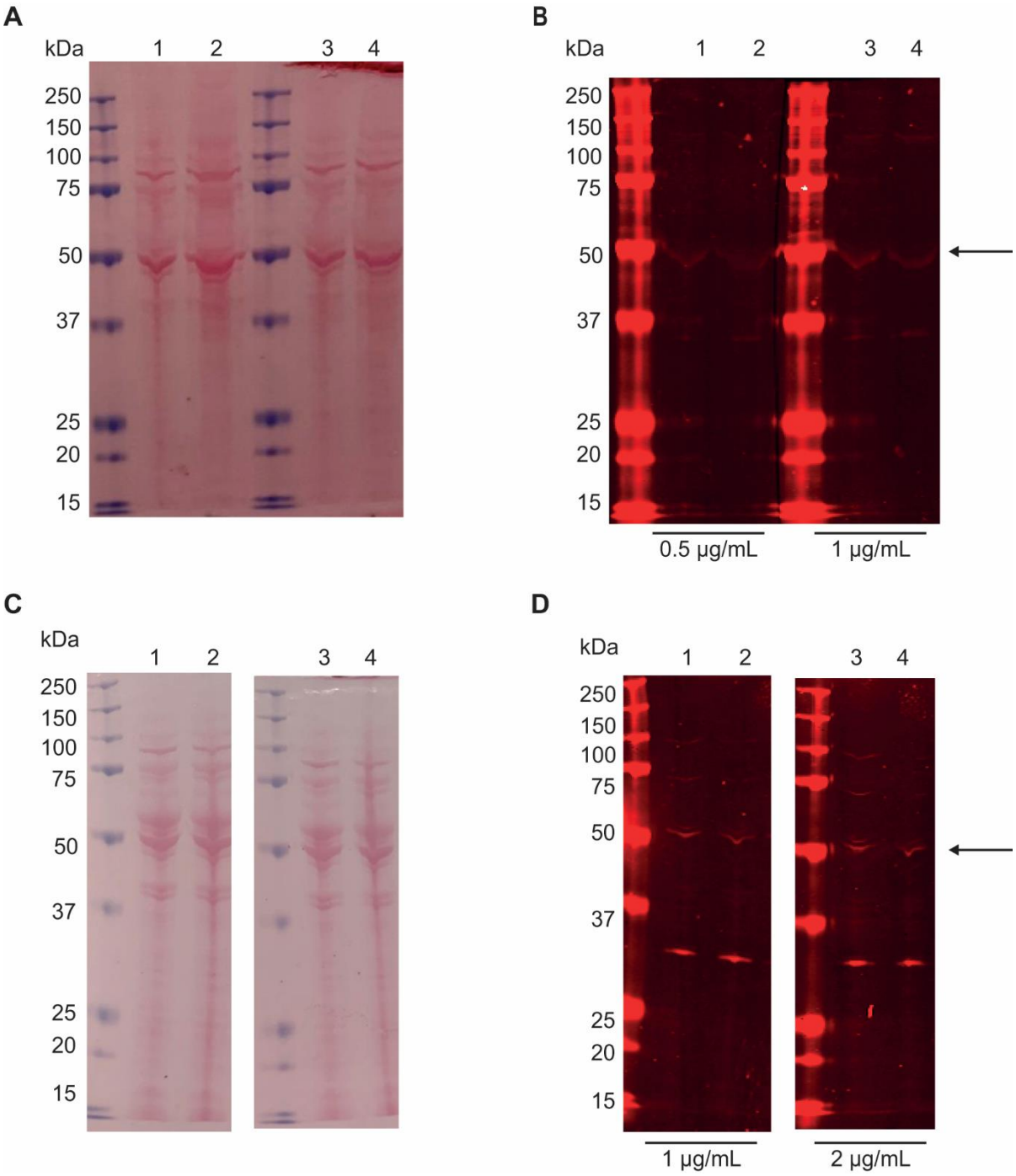

Figure 6.10 Tests for Tb927.11.11750 anti-peptide antibodies.

A and B: Ponceau staining and Western blot using Tb927.11.11750 anti-peptide antibody raised by Davids Biotechnologie. Lane 1 to 4: 1% TX-100 lysates equivalent to 5×10^6 BSF WT cells. The concentration of primary antibody used is indicated at the bottom of panel B. C and D: Ponceau staining and Western blot using Tb927.11.11750 anti-peptide antibody raised by Moravian Biotech. Lanes 1 to 2: 1% TX-100 lysates equivalent to 2.5×10^6 BSF WT cells. Lane

2 and 4: 1% TX-100 lysates equivalent to 2.5 × 10^6 Tb927.11.11750 overexpressing cells. The concentration of primary antibody used is indicated at the bottom of panel D.

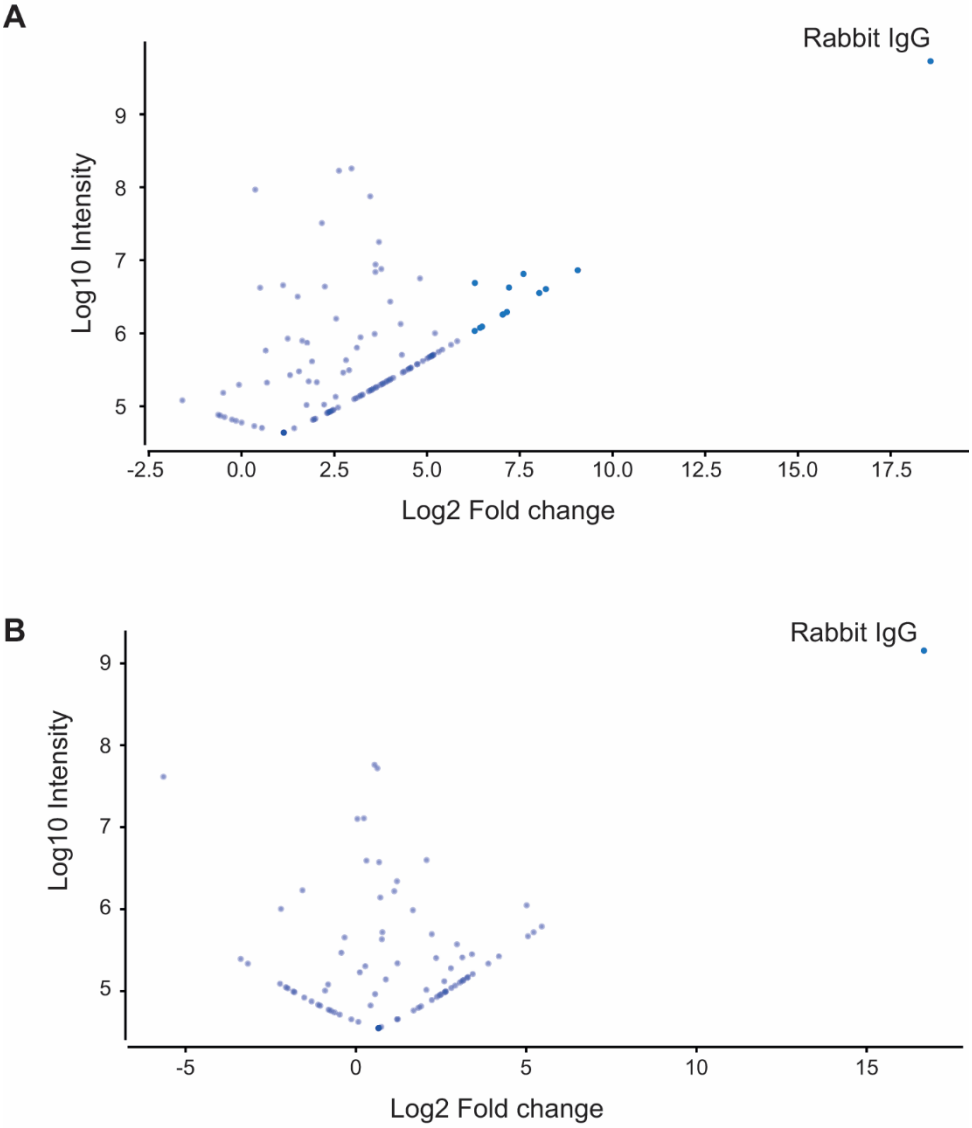

Figure 6.11 Plot of proteins identified from Tb927.11.11750 anti-peptide antibody pull down proteomics experiments.

A and B: 1% TX-100 cell lysates equivalent to 1 × 10^8 BSF WT cells were incubated with (lanes 1 and 3) and without (lanes 2 and 4) 2 µg Tb927.11.11750 anti-peptide antibodies that were raised by Davids Biotechnologie and Moravian Biotech, respectively. Eluted proteins from the protein A beads were separated by denaturing gel and the band regions at ~50 kDa were cut off for proteomics identification. Proteins identified by mass spectrometry were plotted by log10 value of intensity of the peptides (y axis) against log2 fold change between the peptides detected in the Tb927.11.11750 bait (with antibodies) versus control (without antibodies) sample.

6.3.5 Introducing an ectopic copy of internal 12Myc tagged Tb927.11.11750

After the failed attempts at generating useful Tb927.11.11750 specific anti-peptide antibodies, a different strategy for tagging the protein which could facilitate later pull-down experiments was designed. An overexpression plasmid for expressing a 12×c-Myc internal tagged version of Tb927.11.11750 was made using the Gibson assembly method (Gibson *et al.*, 2009), as shown in Fig 6.12A. Basically, an 12×Myc insert was PCR amplified with primers OL19 and OL20 from a synthetic gene sequence in which 21 bp of the sequences encoding the first and last Myc repeat had been re-coded (designed with help of Dr Sam Duncan and synthesised by Genscript). Meanwhile, a vector was also PCR amplified from a previously generated plew100_v5 *Tb927.11.11750* overexpression plasmid (Wirtz *et al.*, 1999) with ends containing the two Myc sequence with different codons using primers OL21 and OL22. After Gibson assembly reaction, the 12×Myc sequence was inserted between the signal peptide and the rest of the *Tb927.11.11750* DNA sequence (the tagged protein is termed SP-12Myc-Tb927.11.11750 hereafter). This construct was then linearised with Not I and transfected into *T. brucei* BSF and the cells were selected with phleomycin. Clones were obtained by limit dilution and the expression of SP-12Myc-Tb929.11.1175 protein from three clones was analysed by anti-Myc Western blot under +/- tetracycline (Tet) conditions. Ponceau staining of the membrane shows similar amount of cells were loaded among different samples (Fig 6.12B); anti-Myc Western blot (Fig 6.12C) of the same membrane shows the detection of bands at ~ 80 kDa (theoretical MW of Tb927.11.11750-12Myc protein is ~68 kDa) in all three tested clones in the + Tet samples but not in - Tet samples, suggesting SP-12Myc-Tb927.11.11750 protein is expressed under Tet control in these three clones.

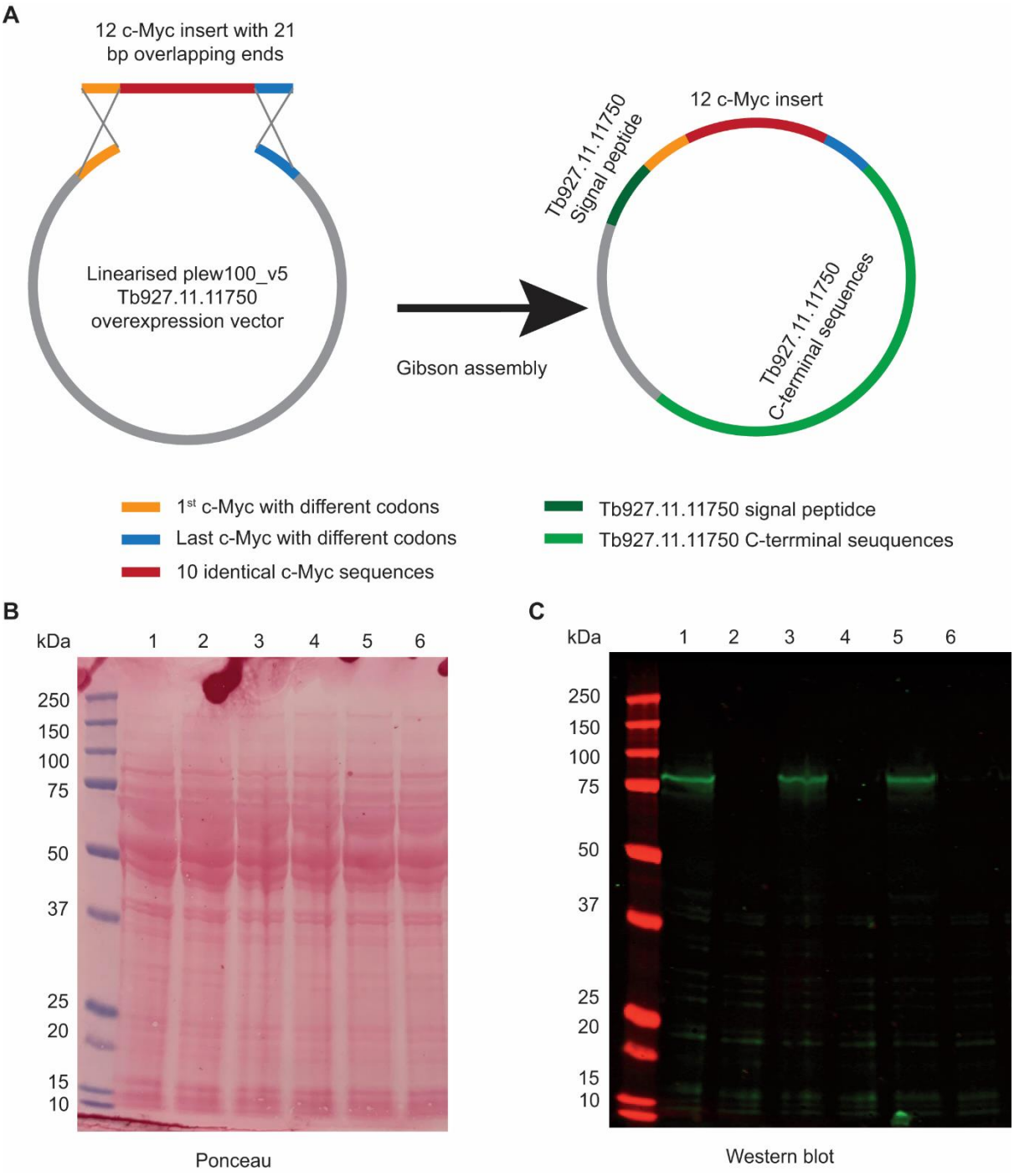

Figure 6.12 Internal 12×Myc tagging of Tb927.11.11750.

A: Scheme of construction of plew100_v5 SP-12Myc-Tb927.11.11750 plasmid using Gibson assembly. Different fragments of the constructs are annotated by colour coding. Yellow: First Myc with different codons. Blue: Last c-Myc with different codons. Red: 10 identical Myc sequences with the same codons. Dark green: Tb927.11.11750 signal peptide sequence. Light green: Tb927.11.11750 C-terminal sequence without signal peptide. B and C: Ponceau staining and anti-Myc Western blot of cell lysates from three clones expressing SP-12Myc-Tb927.11.11750 with or without Tet induction. For each lane, 5 e 6 cells per lane were lysed with SDS sample buffer, subjected to SDS-PAGE and transferred to nitrocellulose membrane.

The blot was probed with anti-Myc antibody. Lane 1, 3, 5 are + Tet samples of clone 1, 2, 3. Lane 2, 4, 6 are - Tet samples of clone 1, 2, 3.

6.3.6 SP-12Myc-Tb927.11.11750 is present in high molecular weight complexes

Previous TbdeAc-3Myc pull down experiments had shown that it forms a ~240 kDa complex. To investigate whether SP-12Myc-Tb927.11.11750 forms a similar complex, pull down experiments of SP-12Myc-Tb927.11.11750 were performed in parallel with TbdeAc2-12Myc. For the generation of a TbdeAc2-12Myc expressing cell line, the 997 bp ORF of TbdeAc2 was amplified with primers OL7 and OL8 and cloned into pNATx^{12MYC} vector (Alsford and Horn, 2008). The plasmid was linearised with AscI and transfected to the BSF WT cells, the 613 bp C-terminal ORF of TbdeAc2 within the construct was used for targeting the origin locus of the gene through homologous recombination. Cells were further selected with BSD and subcloned (Fig 6.13D). Using previously optimised detergent and elution conditions to IP low abundant transmembrane proteins, SP-12Myc-Tb927.11.11750 expressing cells were lysed with 0.5 % digitonin lysis buffer, and immunoprecipitation was performed using Myc-Trap^{TM} beads. The beads were then washed three times and proteins bound to the beads were eluted with SDS sample buffer and analysed by denaturing gel following by anti-Myc Western blot. An identical amount 1% TX-100 SP-12Myc-Tb927.11.11750 cell lysate served as a positive control (Fig 6.13A). Both lysis condition could extract the protein with a comparable efficiency while there is no band detected in wild type lysate sample. For compatible elution conditions for nativePAGE, c-Myc peptide elution was tested after pulling down from a SP-12Myc-Tb927.11.11750 cell lysate (Fig 6.13B). The anti-Myc Western blot shows success using this elution method. Sample replicates were then applied to nativePAGE and analysed by anti-Myc Western blotting, this is accompanied by TbdeAc2-12Myc IP products with the same IP and detection conditions. TbdeAc2-12Myc was detected in complexes at ~ 240 kDa, as before, while SP-12Myc-Tb927.11.11750 was found in multiple complexes ranging from ~180 to 430 kDa (Fig 6.13C). From this analysis, we can conclude that the dominant complexes containing SP-12Myc-Tb927.11.11750 forms are not the same as for TbdeAc2. Nevertheless, some of the SP-12Myc-Tb927.11.11750 does appear with similar native apparent MW as the TbdeAc2-containing complex, and it should be noted that whereas TbdeAc2 was *in-situ* tagged, by necessity, SP-12Myc-Tb927.11.11750 was overexpressed from a tetracycline-inducible ectopic copy.

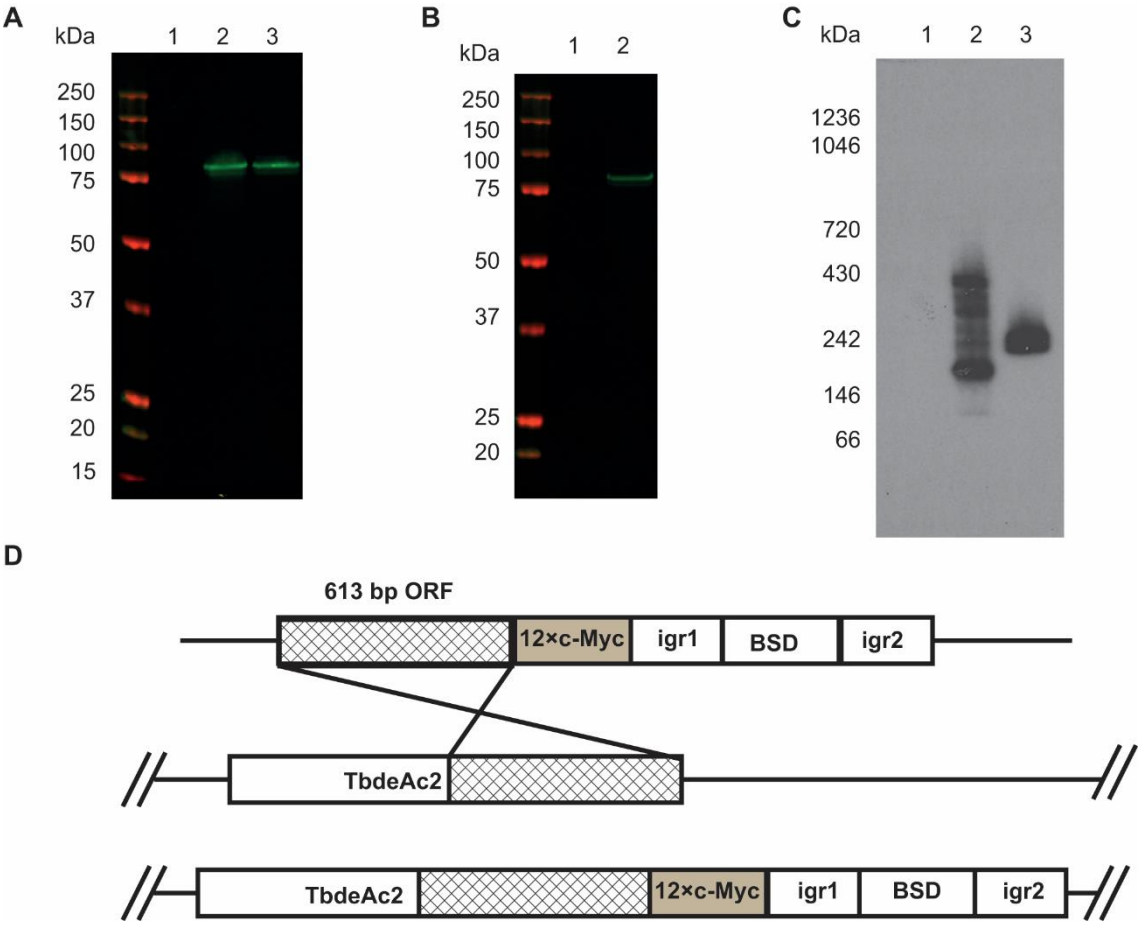

Figure 6.13 Overexpressed SP-12Myc-Tb927.11.11750 forms multiple complexes.
A. Optimisation of detergent conditions for extraction and pull down of SP-12Myc-Tb927.11.11750. Lane 1: BSF WT 1% TX-100 cell lysates, lanes 2 and 3 are SP-12Myc-Tb927.11.11750 expressing BSF 1% TX-100 and 0.5% digitonin cell lysates, respectively. These cell lysates equivalent to 5×10^7 cells per lane were immunoprecipitated with Myc-Trap agarose beads and subjected to denaturing gel further analysed by anti-Myc Western blot. B. Test of c-Myc peptide elution conditions. Myc-Trap pull downs from 0.5% digitonin lysates of BSF WT (lane 1) and Tb927.11.11750-12Myc expressing BSF cells (lane 2) were performed as described before, and proteins were eluted by 2.5 mg /mL c-Myc peptide elution buffer three times. Combined elutes were then analysed by anti-Myc Western blot. C. NativePAGE anti-Myc Western blot of the cMyc peptide eluates of My-Trap pull downs from BSF WT (lane 1), SP-12Myc-Tb927.11.11750 overexpressing (lane 2) and TbdeAc2-12Myc expressing (lane 3) BSF 0.5% digitonin lysates. .D Scheme for *in situ* tagging of TbdeAc2 with 12 c × Myc at C terminus using $pNAT^{12MYC}$ vector. Linearised $pNAT^{12MYC}$ vector containing 631 bp C terminal ORF of TbdeAc2 and 12 c × Myc, blasticidine resistance cassette (BSD) in frame of intergenic region 1(igr1: βα tubulin mRNA) and igr2 (actin mRNA) was transfected to the BSF WT cells, *in situ* tagging of TbdeAc2 was achieved by homologous recombination.

6.3.7 Proteomics identification of **SP-12Myc-Tb927.11.11750 binding partners**

To find out whether TbdeAc2 is a specific partner for SP-12Myc-Tb927.11.11750, immunoprecipitation of lysates from BSF WT and SP-12Myc-Tb927.11.11750 expressing cell lines were performed with the standardised pull-down method established above. Proteins eluted with c-Myc peptide were subjected to NativePAGE, and bands containing SP-12Myc-Tb927.11.11750 complexes on a duplicate gel (the regions were annotated in rectangles in Fig 6.13C) and corresponding bands from the BSF WT sample were excised and sent for in gel trypsin digestion followed by LC-MS/MS analysis. Proteomics data were analysed using MaxQuant software and a newly developed data analysis method written in Python called ProtRank (Medo, Aebersold and Medová, 2019), The protein groups identified were displayed on a plot of the sum iBAQ score of peptides (y-axis) and enrichment rank (x-axis) between the bait versus control. By doing this, bait protein SP-12Myc-Tb927.11.11750 has the highest enrichment and its putative partner protein TbdeAc2 was the second most highly enriched protein (Fig 6.14A). This data strongly supports the physical association of SP-12Myc-Tb927.11.11750 with TbdeAc2.

The data were also processed in a different way that plots the experimental rank (x-axis) against the rank order of estimated abundances of the protein groups, generated from data in (Tinti, Maria Lucia S Güther, *et al.*, 2019), on the y-axis (the higher the protein ranks, the lower abundance it has) (Fig 6.13B). This plot also shows a clear association between Tb927.11.11750 and TbdeAc2 (Tb927.3.2610), and also emphasises that both are very low abundance proteins in the total trypanosome proteome.

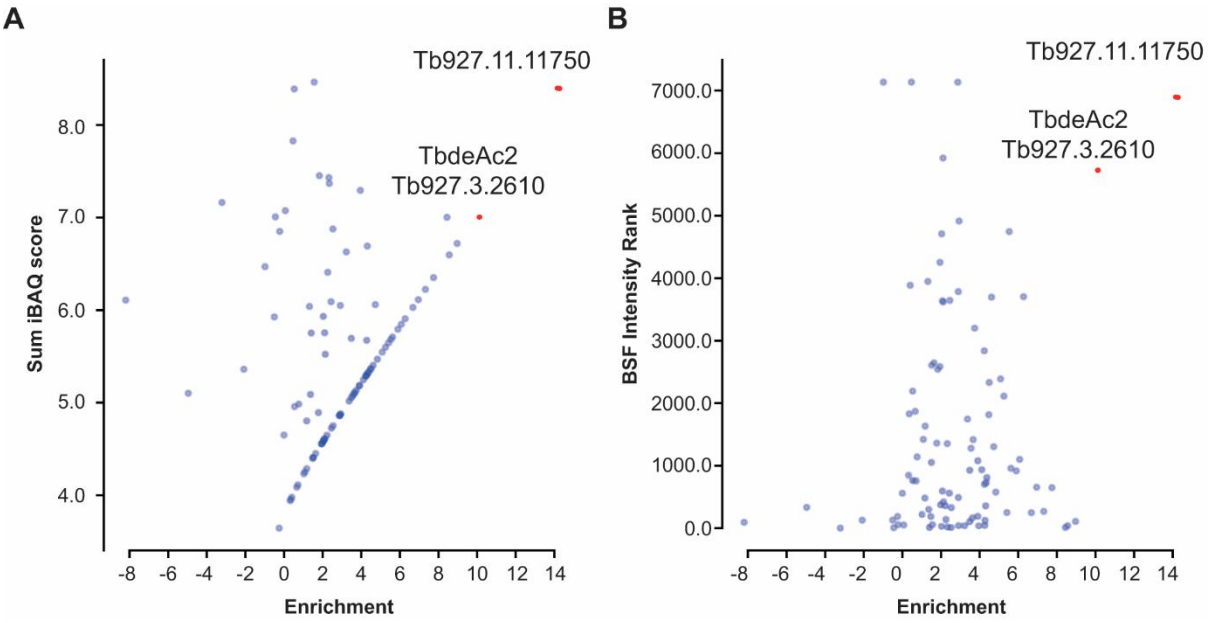

Figure 6.14 Identification of TbdeAc2 by co-immunoprecipitation with SP-12Myc-Tb927.11.11750 from BSF *T. brucei* digitonin lysates.

A. A plot comparing protein groups present in the anti-cMyc immunoprecipitates from SP-12Myc-Tb927.11.11750 expressing cell lysates versus BSF WT cell lysates. Each protein group is plotted according to its Sum iBAQ values (y-axis), calculated by MaxQuant, and the enrichment rank (x-axis). The enrichment rank was computed with the ProtRank algorithm using the iBAQ values calculated by MaxQuant. Tb927.11.11750 and Tb927.3.2610 are highlighted in red. (B) Relative intensity plot using a new algorithm. The same data as (A) are plotted with a different y-axis, whereby each protein group is assigned an intensity rank from the most abundant protein group (1) to least abundant protein groups (7,149) based on their summed eXtracted Ion Currents (XICs) for the total BSF proteome.

While the reciprocal pull-downs (of Tb927.11.11750 with TbdeAc2-3Myc and TbdeAc2 with SP-12Myc-Tb927.11.11750) are consistent with the two proteins forming a stable complex, the differences in their nativePAGE patterns suggests that overexpressed SP-12Myc-Tb927.11.11750 is not exclusively co-localised with TbdeAc2. The proteomics data in (Fig 6.14A) show that the overexpressed SP-12Myc-Tb927.11.11750 protein is indeed present in excess over TbdeAc2 in the Myc-Trap pull down according to the summed iBAQ scores. This is in contrast to the native situation, where the protein rankings for TbdeAc2 and Tb927.11.11750 and Tb927.11.11740 together are 5715th and 6882nd out of 7149 protein groups, respectively (Ji et al., 2021). These protein rankings, and the apparent homogeneity of the ~240 kDa TbdeAc2-3Myc and TbdeAc2-12Myc *in-situ* tagged complexes, as visualised by native PAGE (Fig 6.1C, lane 2 and Fig 6.13C, lane 3), suggest that native TbdeAc2 and Tb927.11.11750 are perhaps present in 2 : 1 ratio in a bimolecular complex.

6.4 TbdeAc2 and Tb927.11.11750 are localised in endoplasmic reticulum

Previous work has shown that GPI anchor precursors are present in the endoplasmic reticulum (ER) (Vidugiriene and Menon, 1994). It is therefore a reasonable assumption that the enzymes catalysing GPI synthesis also reside in ER. Nevertheless, apart from the ER-localisation of TbdeAc1(Güther *et al*., 2003) and very recently for GPI2 (Jenni *et al.*, 2021), there are no experimental data on the localisation of GPI anchor biosynthesis pathway enzymes. To address this deficit, the subcellular localisations of *in situ* tagged TbdeAc2-12Myc and SP-12Myc-Tb927.11.11750 were analysed in *T. brucei* BSF cells by immunofluorescent microscopy (IFM). TbdeAc2-12Myc (Fig 6.15A) and SP-12Myc-Tb927.11.11750 (Fig 6.15C) were found to substantially co-localise with the known ER protein marker BiP (Bangs *et al.*, 1993). Both stains showed typical ER reticulated staining and clear perinuclear rings. This is not only consistent with TbdeAc2 being an ER resident membrane enzyme, as expected, but also demonstrates the co localisation of TbdeAc2 and Tb927.11.11750, providing supporting evidence that Tb927.11.11750 may be involved in the GPI anchor biosynthesis pathway. The absence of anti-Myc staining in the BSF WT cells demonstrates the specificity of the staining for transgenic TbdeAc2-12Myc and SP-12Myc-Tb927.11.11750.

We also tried to localise *in situ* tagged Tb927.11.11740-3HA, but the signal was below the limit of detection (data not shown). The TrypTag.org detection for Tb927.11.11740-mNeonGreen (Fig 6.15B) was inconclusive with possible suggested localisations of “cytoplasm, nuclear lumen and endocytic” (Dean *et al*., 2017). In any case, the pattern of Tb927.11.11740-mNeonGreen staining suggests it is unlikely to be a ER resident protein.

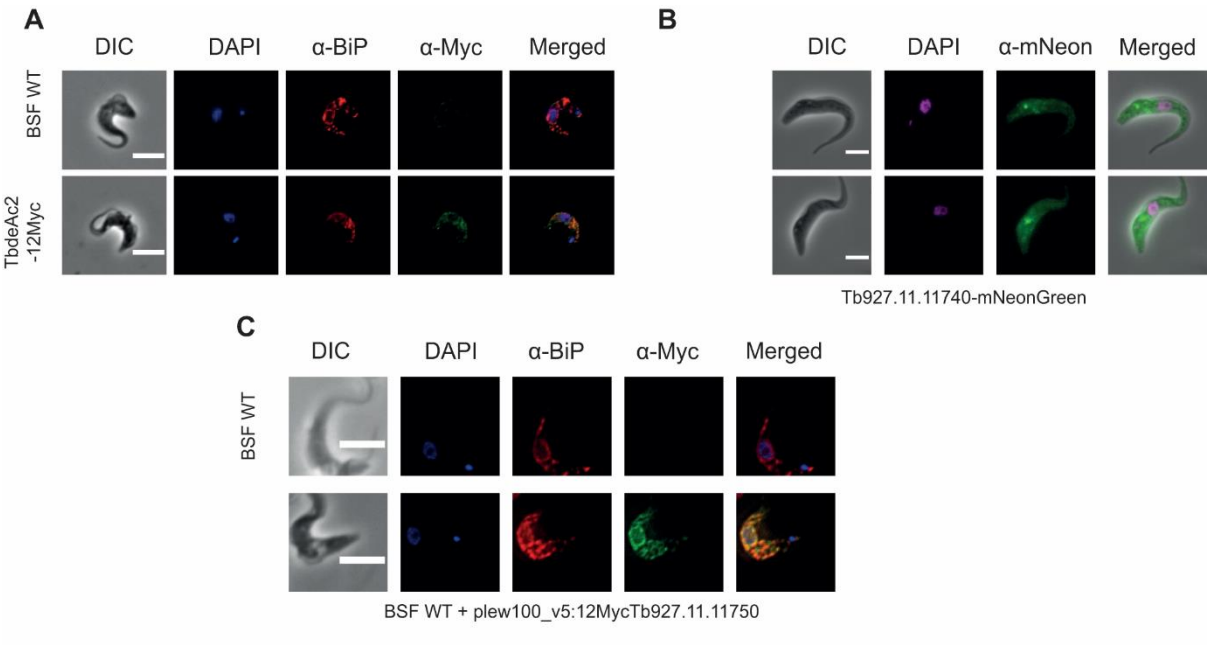

Figure 6.15 Localisations of TbdeAc2-12Myc, Tb927.11.11740-mNeonGreen and SP-12Myc-Tb927.11.11750.

A and C: Fixed and permeabilized BSF WT *T. brucei* and parasites expressing *in situ* tagged TbdeAc2-12Myc and internal 12Myc tagged SP-12Myc-Tb927.11.11750 were co-stained with anti-Myc antibodies to detect TbdeAc2/Tb927.11.11750 (green) and anti-BiP to detect the ER (red). The cells were also stained with DAPI to detect the nuclear and kinetoplast DNA. The parasites were further imaged by DIC. Similar staining patterns were observed in the anti-BiP and anti-Myc channels suggesting both TbdeAc2 and Tb927.11.11750 localise in the ER. B: Localisation of *in situ* C-terminal mNeonGreen tagged Tb927.11.11740 (two fields) was analysed by TrypTag.org (Dean, Sunter and Wheeler, 2017). The pattern of mNeonGreen signal suggested that Tb927.11.11740 was unlikely an ER resident protein. Scale bars: 5 µm.

6.5 Knock down of Tb927.11.11740 and Tb927.11.11750

Having confirmed that TbdeAc2 forms a protein complex with Tb927.11.11750 but not with Tb927.11.11740, and provided evidence of them being ER resident proteins, the next question we wanted to address was: What are the roles that Tb927.11.11750 and/or Tb927.11.11740 might play in the GPI anchor biosynthesis pathway? Due to the high sequence identities between these two genes and encoded proteins, several different gene manipulation approaches were applied to *Tb927.11.11750* and *Tb927.11.11740* simultaneously and individually to try to generate different types of mutant cell lines for functional studies. By comparing the GPI anchor biosynthesis pathway products between the mutant cell lines and WT cell lines, we hoped that we might uncover the functions of Tb927.11.11750 and/or Tb927.11.11740.

6.5.1 Knock down of *Tb927.11.11740* and *Tb927.11.11750* using a p2T7 RNAi construct

Since Tb927.11.11740 and Tb927.11.11750 have the same predicted catalytic domains and motifs (Fig 6.6B) but different subcellular localisations (Fig 6.15), we hypothesize that these two proteins have similar enzymatic functions but in different subcellular environments. To preclude functional compensations that may happen between Tb927.11.11740 and Tb927.11.11750, a knock down construct that targets both genes was assembled in the p2T7 plasmid (Alibu *et al.*, 2005) to generate inducible RNAi strains with primers OL9 and OL10. The sequence that was chosen for the knock down construct was determined through RNAit tool (Redmond, Vadivelu and Field, 2003) in which the sequences were BLAST searched against the whole *T. brucei* genome to minimise off target effects. Six clones were picked after transfection and hygromycin selection for further analysis. The growth curves of these six clones were monitored with and without RNAi induction by tetracycline (Tet) over 7 days. There was no significant growth defect associated with the knock down (Fig 6.16A), suggesting that both genes are dispensable for BSF *T. brucei* survival. However, when the extent of the RNAi knockdown was measured by real time qRT-PCR (Fig 6.16B) with primers OL35 and OL36 a reduction of only ~40% of the targeted mRNAs was observed after Tet reduction for 8 h and this reduced further to ~20% when after 24 h of Tet induction. This modest extent of RNAi knockdown was insufficient for biochemical characterisation and so this RNAi knockdown approach was abandoned.

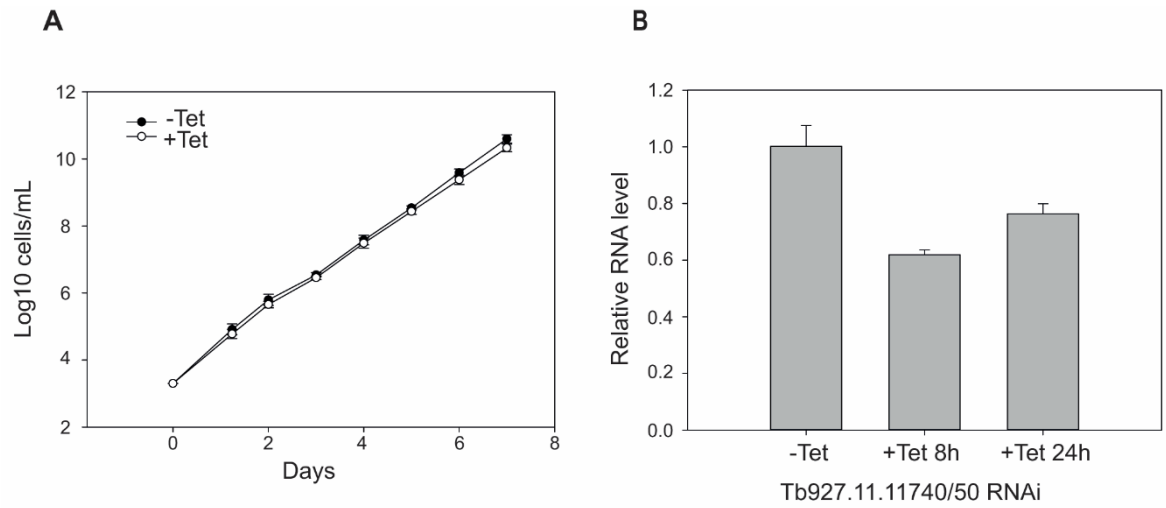

Figure 6.16 Knock down of *Tb927.11.11740* and *Tb927.11.11750* using p2T7 construct.
A: Cumulative growth of six clones following tetracycline (Tet) induced knockdown of *Tb927.11.11740* and *Tb927.11.11750*. Minus Tet: solid line with black solid dots; Plus Tet:

solid line with hollow dots. B: Real time quantitative RT-PCR analysis of the relative abundance of *Tb927.11.11740* and *Tb927.11.11750* mRNA transcripts after 8 h and 24 h of RNAi induction with tetracycline (Tet) of clone 1. The y axis represents the relative abundance mRNA of 8h and 24h time point +Tet sample compared to that of the RNA from -Tet cells (set at 1.0). Error bars indicate one standard deviation.

6.5.2 Knock down *Tb927.11.11740* and *Tb927.11.11750* through pRPa RNAi construct

The inefficient knock down using the p2T7 RNAi construct led us to try an alternative RNAi method based on pRPa plasmid which was established to generate stem loop RNAi transcripts from a specific ribosomal RNA (rRNA) locus in *T.brucei* (Alsford and Horn, 2008). To enable the construct to integrate into this specific targeted rRNA locus, BSF 2T1 WT cell line was used as the parental cell line (Alsford *et al.*, 2005). The target sequence for RNAi construct was the same as p2T7 construct which was a 486 bp identical region of *Tb927.11.11740* and *Tb927.11.11750*, amplified with primers OL11 and OL12. Three independent RNAi knockdown clones were isolated and induced with tetracycline. No significant growth defect was associated with Tet induced knockdown (Fig 6.17A). The extent of the RNAi knockdown level of these three clones was measured by real time qRT-PCR after 24 h Tet induction (Fig 6.17B). Among these three clones, clone 1 had the highest knockdown level with a reduction of ~70% while clone 2 and 3 had ~60% reduction. To further investigate the knockdown efficiency of clone 1, a time course of knockdown was examined at 8, 24, 48 and 72 h time points (Fig 6.17C). The RT-qPCR assay revealed that the most efficient knock down for *Tb927.11.11740* and *Tb927.11.11750* was reached after 24 h with a ~70% reduction at the mRNA level and this reduced to ~ 50% after 72 h of induction. While the knockdown was not as deep as we would have liked, we decided to perform functional studies on this RNAi cell line to check whether there were any observable differences in the GPI anchor biosynthesis between the RNAi and WT cell lines.

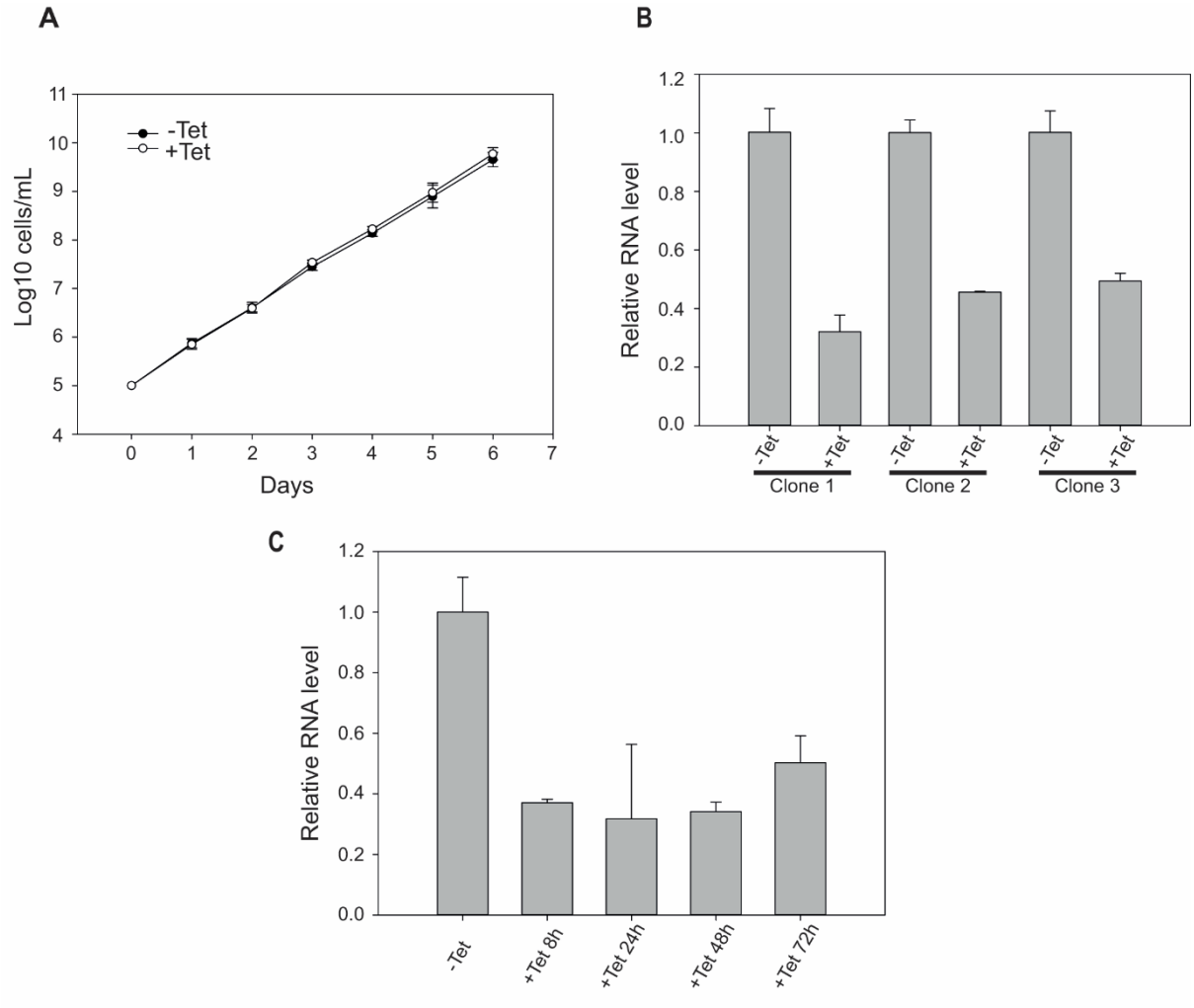

Figure 6.17 Knock down of *Tb927.11.11740* and *Tb927.11.11750* using pRPa construct. A: Cumulative growth of three clones following tetracycline (Tet) induced knockdown of *Tb927.11.11740* and *Tb927.11.11750*. Minus Tet: solid line with black solid dots; Plus Tet: solid line with hollow dots. B: Real time quantitative RT-PCR analysis of the relative abundance of *Tb927.11.11740* and *Tb927.11.11750* mRNA transcripts after 24 h of RNAi induction with tetracycline (Tet) of clone 1, 2 and 3. The y axis represents the relative abundance of each mRNA compared to that of the RNA from -Tet cells (set at 1.0). Error bars indicate one standard deviation. C: Time course of real time qRT-PCR analysis of relative mRNA abundance of *Tb927.11.11740* and *Tb927.11.11750* transcripts of RNAi knock down clone 1 after 8 h, 24 h, 48 h and 72 h Tet inductions.

6.5.3 Comparison of the GPI products synthesized by the trypanosome cell-free system between *Tb927.11.11740* and *Tb927.11.11750* RNAi and BSF 2T1 WT cell lines

As Tb927.11.11750 is closely associated with TbdeAc2 and both Tb927.11.11750 and Tb927.11.11740 contain a predicted lipase motif GXGXG, one a reasonable hypothesis

for their roles in GPI anchor biosynthesis pathway is that one or other might be the unknown component GPI phospholipase A2 (PLA2), which is the enzyme catalysing the next step after TbdeAc2.

The different subcellular localisations of Tb927.11.11740 and Tb927.11.11750 could also be explained as a previous study showed that, in addition to GPI fatty acid remodelling in the ER (Masterson *et al.*, 1990) here is a separate acyl chain remodelling for mature VSG-linked GPI anchors in endocytic pathway (Buxbaum *et al.*, 1994).

To test this GPI PLA2 hypothesis, three *T. brucei* cell-free systems were prepared of from BSF 2T1 WT cells and from the *Tb927.11.11740* and *Tb927.11.11750* RNAi knockdown line harvested before and after Tet uninduction to perform the radiolabelling of GPI glycolipid intermediates in the GPI anchor biosynthetic pathway, as described previously (Güther *et al.*, 1994) (Fig. 6.18A). Briefly, the cells were subjected to hypotonic lysis and the membrane fraction (cell-free system) containing the GPI biosynthetic machinery was recovered by centrifugation. Pulse-labelling of the cell-free system with GDP-[^{3}H]Man was performed and the labelled glycolipids were extracted and resolved by high-performance thin layer chromatography (HPTLC) and visualised by auto-fluorography. In this experiment, BSF 2T1 WT cell line (Fig 6.18B, lane 1) sample was served as positive control, and the pattern of the labelled GPI intermediates was the same when using cell-free system prepared from *T. brucei* (variant MITatl.4) BSFs that were isolated from infected rats (Güther, Masterson and Ferguson, 1994) . The *Tb927.11.11740* and *Tb927.11.11750* RNAi knockdown cell lines were collected after 24 h induction with or without Tet for cell-free system preparation. The radiolabelled GPI anchor products on the HPTLC plate are tentatively assigned according to their migration positions. If the PLA2 hypothesis was correct, we might expect that there would be less glycolipid θ product in the induced RNAi knockdown cell line. However, there were no significant differences in the labelled GPI anchor intermediates among the three tested cell-free systems, as shown in (Fig 6.18B). This result suggested that either Tb927.11.11740 and Tb927.11.11750 do not function as a GPI PLA2 enzymes, or the 70% RNAi knockdown is insufficient for detection of any functional perturbation. Therefore, we determined that a null or conditional null mutant cell line for *Tb927.11.11740* and *Tb927.11.11750* was necessary to provide a clean and stable genetic background to perform functional studies.

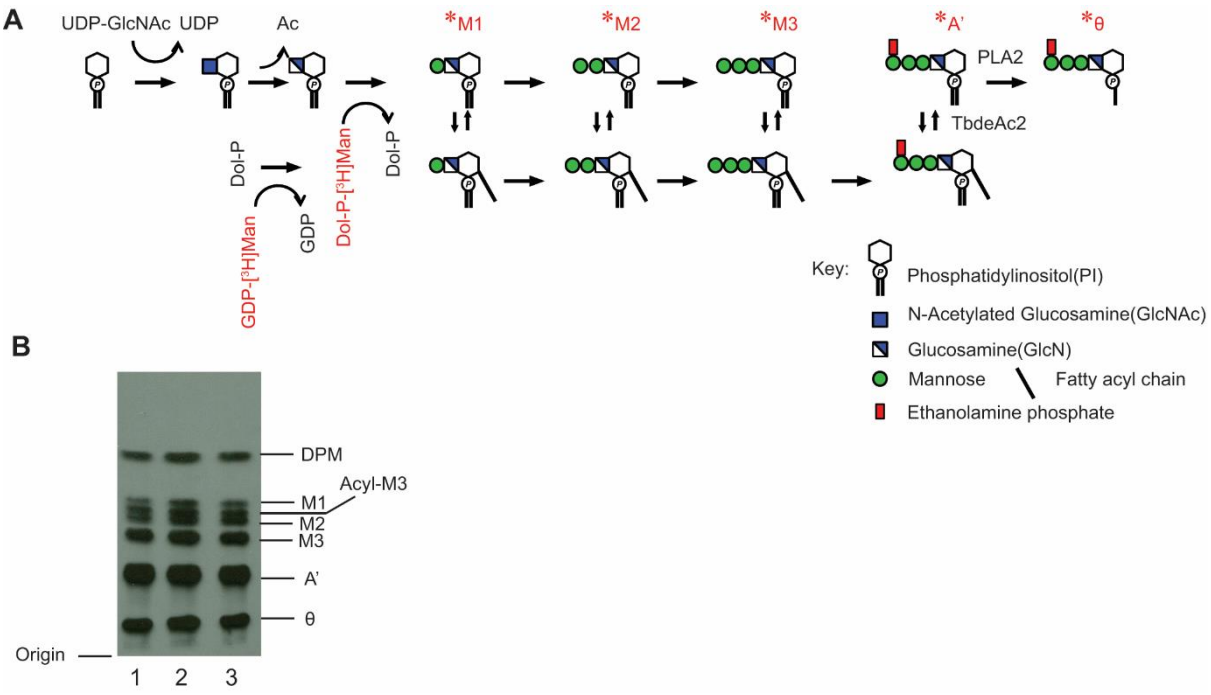

Figure 6.18 GPI anchor products in BSF 2T1 WT and *Tb927.11.11740* & *Tb927.11.11750* RNAi knockdown *T. brucei* cell-free system.

A. Scheme of cell-free radiolabelling system in *T.brucei*. The cell-free system is incubated with GDP-[^{3}H]Man in the presence of 1 mM UDP-GlcNAc. All the radiolabelled glycolipid intermediates formed in such a reaction are annotated in red with asterisks. The key for the components of the glycolipids were indicated at the bottom right of the scheme. B. *T. brucei* cell-free system was prepared for BSF 2T1 (lane 1) and *Tb927.11.11740* & *Tb927.11.11750* RNAi knockdown Tet uninduced (lane 2) and induced (lane 3) for 24 h. Glycolipids of each sample (equivalent to 2.5 × 10^6 cells/lane) were extracted with solvent and analysed by HPTLC and fluorography. The descriptors θ, A', C, M3, M2, acyl-M3 and M1 next to lane 3 correspond to glycolipids θ, A', C, Man_3GlcN-PI, Man_2GlcN-PI, Man_3GlcN(acyl)PI and Man_1GlcN-PI respectively. DPM corresponds to dolichol phosphate-mannose.

6.6 Attempts at CRISPR-Cas9 mediated gene knock out of *Tb927.11.11740* and *Tb927.11.11750* in BSF *T. brucei*

6.6.1 Attempt at the generation of *Tb927.11.11740* and *Tb927.11.11750* null mutant by CRISPR-Cas9

The RNAi knock down experiments of *Tb927.11.11740* and *Tb927.11.11750* suggested that both genes might be non-essential in BSF *T. brucei*, however the maximal knockdown of transcripts was only ~70% based on qRT-PCR analysis so that this conclusion is questionable. Nevertheless, to explore the function of these two proteins, CRSIPR Cas9 mediated gene editing was attempted, aiming to knock out both genes

simultaneously through a previously described strategy (Rico *et al.*, 2018). Basically, an sgRNA sequence was selected before a PAM motif in the identical regions of both genes, using BLAST searches to select sequences likely to minimise potential off target effects. The sgRNA construct was amplified with primers OL39 and OL40 and, along with a repair template with a puromycin acetyl transferase resistance cassette (PAC), was transfected into *T. brucei* BSF expressing Cas9 under Tet regulation (described as "Cas9 WT" hereafter). After tetracycline-induced Cas9 editing, the cells were selected with puromycin and sub-cloned for genotype screening (Fig 6.19A). We hoped that Cas9 editing would lead to double strand DNA breaks within both alleles of *Tb927.11.11740* and *Tb927.11.11750* and that the whole locus, including the intergenic region (igr), would be replaced by a PAC resistance cassette through homologous recombination (Fig 6.19B). The repair template used in this experiment was PCR amplified from a plasmid containing PAC flanked with 800 bp 5'UTR sequence of *Tb927.11.11740* and 500 bp 3'UTR sequence of *Tb927.11.11750* using primers OL23 and OL28. The template was first assembled to pUC57 vector, primers OL23 and OL24 were used for amplifying the 800 bp 5' UTR of *Tb927.11.11740*, OL25 and OL26 were used for amplifying the PAC resistance cassette, OL27 and OL28 were used for amplifying 500 bp 3'UTR of *Tb927.11.11750*, OL 29 and OL30 were used for amplifying linearised pUC57 vector.

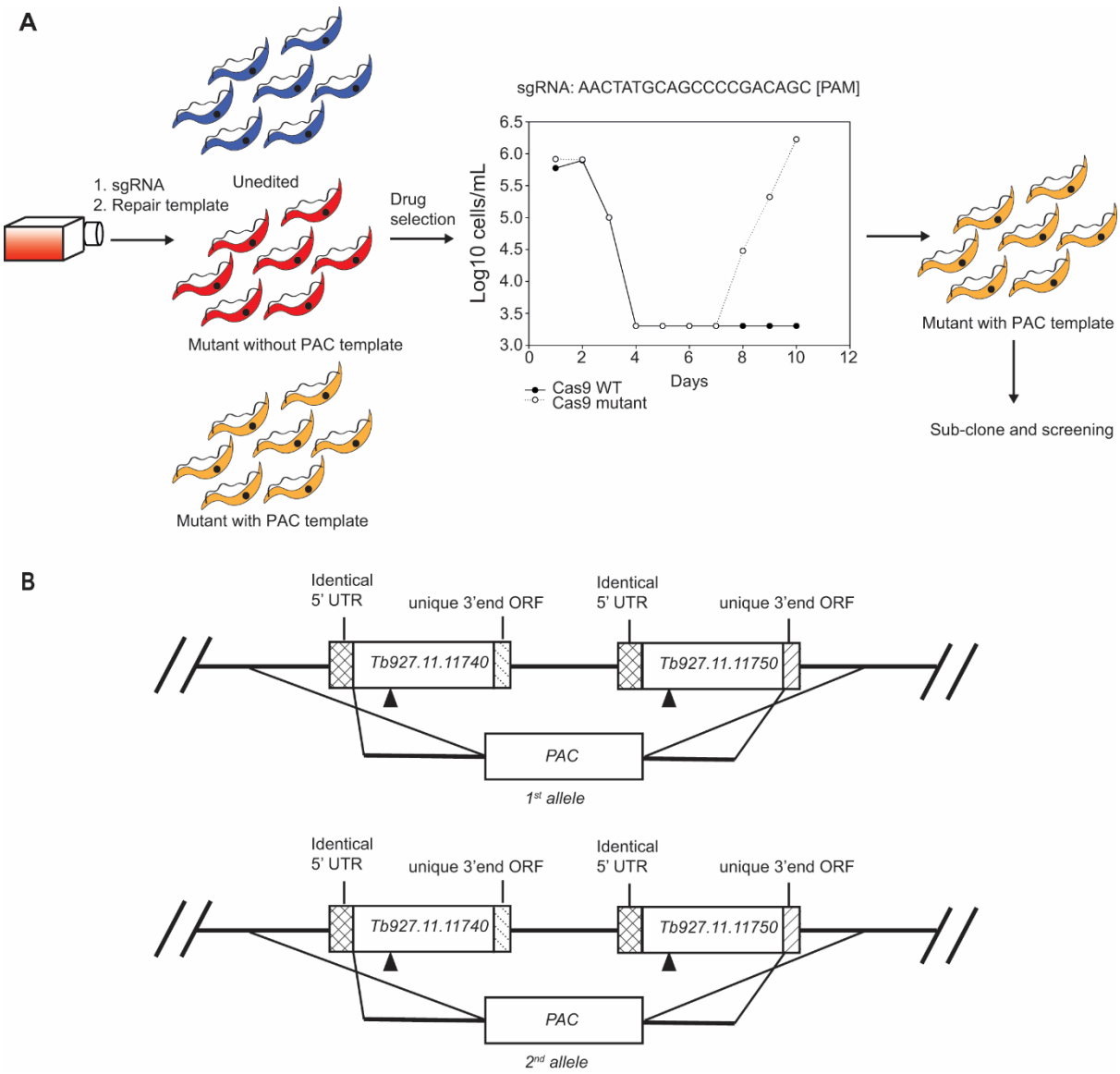

Figure 6.19 Scheme of CRISPR-Cas9 gene editing.

A: A *Tb927.11.11740* and *Tb927.11.11750*-specific sgRNA construct and a repair template were transfected into a BSF *T.brucei* Cas9 WT cell line generated previously to express Cas9 under Tet induction (Rico *et al.*, 2018). After 24 h of Cas9 editing, there should be three cell populations: non-edited cells (blue), edited cells without repair template (red) and edited cells with the puromycin resistance (PAC) template (yellow). Puromycin was then added for selection and the cells recovered after 7 days treatment are subcloned. B: Expected gene homologous recombination after Cas9 editing. Double strand DNA break caused by Cas9 occurs at specific regions (dark triangles) guided by sgRNA. The PAC repair template containing the specific 5'UTR of *Tb927.11.11740* and 3'UTR of *Tb927.11.11750* should integrate into the genome as indicated if the proteins encoded by Tb927.11.11740 and Tb927.11.11750 are non-essential.

6.6.2 Genotyping CRISP-Cas9 edited *Tb927.11.11740* and *Tb927.11.11750* mutants

To check how the repair templates were integrated to Cas9 edited clones, a Southern blot was performed to genotype several clones of the recovered Cas9 mutant cell line. NcoI was selected to perform endonuclease reaction on genomic DNA that was extracted from the cloned cell lines, as described in Materials and Methods. The sequence of the locus gives different predicted restriction fragments sizes for *Tb927.11.11740* (5725 bp), *Tb927.11.11750* (3302 bp) and for the expected PAC-containing gene replacement product (4681 bp), enabling us to distinguish these two genes at the endogenous locus from the locus replaced with PAC (Fig 6.20A). Three Cas9 edited clones were tested along with Cas9 WT cells using digoxigenin (DIG) labelled probes for detection of a common ORF fragment for both genes (the gene of interest, GOI probe, and for the PAC resistance cassette, PAC probe). Primers OL41 and OL42 were used to amplify the GOI probe from the plew100_v1 *Tb927.11.11750* overexpression plasmid which contains the full ORF of Tb927.11.11750 (the construction of this plasmid was introduced in section 6.8). Primers OL43 and OL44 were used to amplify the PAC probe from a plasmid containing full PAC ORF generated in our lab. If the Cas9 mediated gene editing occurred as expected for non-essential genes, the GOI probe should give a two-band restriction fragment pattern for the Cas9 WT cells and no band in the Cas9-induced null mutant cell lines. Conversely, the PAC probe should detect one band in the Cas9 edited clones and no band in the Cas9 wild type cells.

The Southern blot (Fig 6.20C) showed that the PAC repair template had integrated into one genomic locus in Cas9 mutant clone 1 and into two loci in Cas9 mutant clones 2 and 3 (Fig 6.20C). However, the band detected in clone 1 was around 6.5 kb and the bands detected in clones 2 and 3 were 6.5 kb and 4.0 kb. The 6.5 kb and 4.0 kb bands are larger and smaller, respectively, than the predicted 4681 bp fragment, suggesting that PAC integration was not into the *Tb927.11.11740* / *Tb927.11.11750* locus. Further, although the GOI probe blot (Fig 6.20B GOI) showed disappearance of two wild type GOI bands of the wild type cell, these are replaced by either one (clone 1) or two (clones 2 and 3) bands of different sizes. Thus, we concluded that all clones survived by both acquiring PAC *and* by retaining copies of *Tb927.11.11740* and/or *Tb927.11.11750*. This suggested that *Tb927.11.11740* and/or *Tb927.11.11750* are most likely essential genes. Interestingly, in all three clones it is possible that the upper 6.5 kb restriction

fragment may contain the PAC gene together with all or part of *Tb927.11.11740* and/or *Tb927.11.11750*.

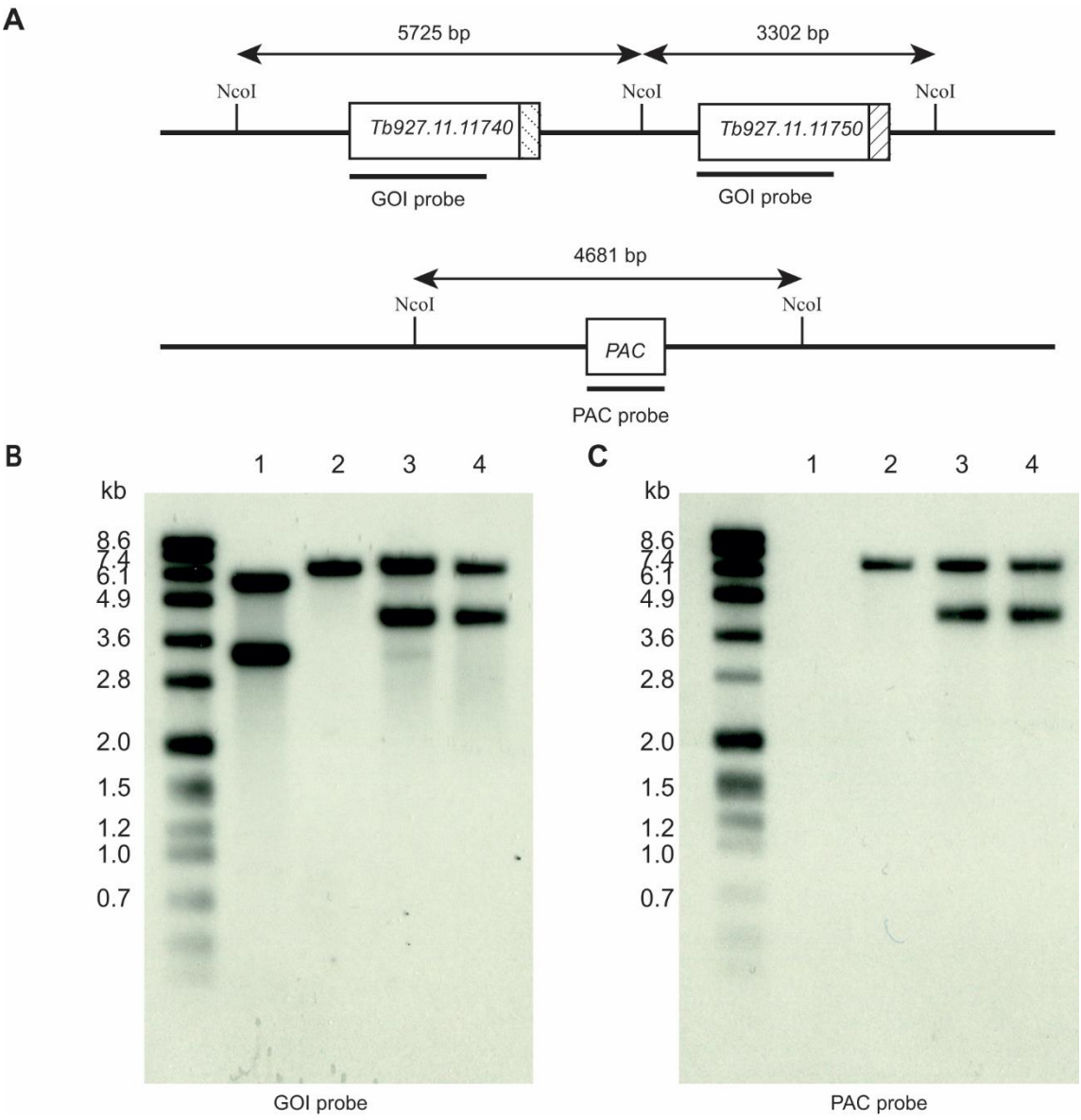

Figure 6.20 CRISPR Cas9 edited clones show aberrant gene rearrangements.

A: Schematic representation of NcoI digestion sites and the expected restriction fragments for detection by the *Tb927.11.11740* and *Tb927.11.11750* GOI probe before and after CRISPR Cas9 editing, assuming both genes to be non-essential. The probes used for detection are indicated below the scheme of the gene locus, the expected sizes of bands detected by the probes are indicated above the scheme of the gene locus. B and C: Southern blot of BSF *T. brucei* gDNA (5 µg/ lane) of Cas9 WT and Cas9 mutant clones 1 to 3 digested with NcoI. GOI ORF (corresponding to 1000 bp identical region of *Tb927.11.11740* and *Tb927.11.11750*) (B) and PAC probe (C) was used for probing after stripping the GOI probe.

6.6.3 Phenotyping CRISPR Cas9 mutant clone 1

To assess whether these aberrant gene rearrangements in the CRISPR Cas9 mutants were due to the essentiality of *Tb927.11.11740* and/or *Tb927.11.11950*, we decided to find out how the PAC repair template was integrated into the genome in the Cas9 mutant clone 1 in which the genotype appeared the simplest among the three mutant clones. The PAC-containing region was amplified and cloned into a plasmid as indicated in (Fig 6.21A). After transfecting this vector to bacteria, 8 clones were picked and sent for plasmid insert sequencing. The sequencing results suggested that there was allelic variation in PAC integration in Cas9 mutant clone 1. Thus, both alleles show replacement of *Tb927.11740* and the intergenic region (igr) with the full-length sequence of the PAC repair template followed by a substantial remnant of *Tb927.11750* after the PAM motif. The sequencing results for both alleles are shown in (Fig 6.21A): The first allele retained 16 bp from the repair template that contains a potential start Met codon, followed by the sequence AQWR, with the *Tb927.11750* remnant encoding a (Δ1-64)Tb927.11.11750 protein in frame with that sequence. The second allele retained 15 bp from the repair template such that the potential Met start site is followed by the sequence AQWQ and an out of frame nonsense version of the *Tb927.11.11750* remnant (there are nonsense codons downstream of the translations shown in Fig 6.21A). Although the genotype of CRSIPR Cas9 mutant clone 1 is complex (i.e., null for *Tb927.11.11740* and containing a single translation-competent truncated version of *Tb927.11.11750*), a cell-free system of this mutant cell line was prepared together with two WT cell line cell free systems. These were used to assess whether CRSIPR Cas9 mutant clone 1 showed any defects in GPI anchor biosynthesis, particularly any deficiency in the production of glycolipid θ products based the hypothesis that Tb927.11.11740 and/or Tb927.11.11750 might function as GPI PLA2 enzymes. The cell-free system preparation and radiolabelling experiments were carried out as described in section 6.5.3 with the addition of a subsequent chase with or without unlabelled excess GDP-Man to push the cell-free system towards the synthesis of the glycolipid θ product. The labelled glycolipids were extracted and resolved by HPTLC and visualised by auto-fluorography. Under these conditions, the Cas9 WT sample formed labelled Man_{1-3}GlcN-PI and glycolipids A' and θ, with the label in the Man_{1-3}GlcN-PI intermediates chasing into glycolipids A' and θ when a cold-chase with GDP-Man was applied (Fig 6.21B, lanes 1 and 2). Any significant diminution in PLA2 activity would be expected to lead to the disappearance or reduction in the intensity of

glycolipid θ and accumulation of label in glycolipid A'. As can be seen (Fig 6.21B, lanes 3 and 4), the CRSIPR Cas9 mutant clone 1 mutant cell-free system behaved identically to the Cas9 WT cell-free system, suggesting no effect on GPI PLA2 activity. In this experiment the samples prepared from *T. brucei* BSF WT cell-free system (Fig 6.21B lanes 5 and 6) were run in parallel with no differences detected in the labelled glycolipids compared to the Cas9 WT cell-free system suggesting that the genome modifications in the Cas9 WT cell line did not interfere with the GPI anchor biosynthesis pathway.

This result leaves us with the question as to whether the absence in GPI biosynthesis deficiencies in the Cas9 mutant clone 1 is due to the fact that neither *Tb927.11.11740* nor *Tb927.11.11750* encode GPI PLA2 activity, or whether the Δ1-64 truncated version Tb927.11.11750 can still encode GPI PLA2 activity, despite lacking a signal peptide. To check whether the truncated *Tb927.11.11750* gene is transcribed, and the therefore could potentially retain function in clone 1, real time qRT-PCR was employed to analyse the mRNA level of *Tb927.11.11750* in mutant clone 1 in comparison with the Cas9 WT cell line. Almost same amount of mRNA was transcribed in both cases (Fig 6.21C) indicating that the Cas9 mutant clone 1 was not a null mutant cell line for *Tb927.11.11750*. Thus, unfortunately, we could not conclude from the cell-free system data whether *Tb927.11.11750* encodes GPI PLA2 activity or not.

In summary, our interpretations of these data are: (i) That *Tb927.11740* and the igr are dispensable to the parasite. (ii) That *Tb927.11.11750* is essential. (iii) That although the predicted MAQWR-(Δ1-64)Tb927.11.11750 fusion protein lacks a signal peptide, sufficient Tb927.11.11750 function is retained for Cas9 mutant clone 1 to survive.

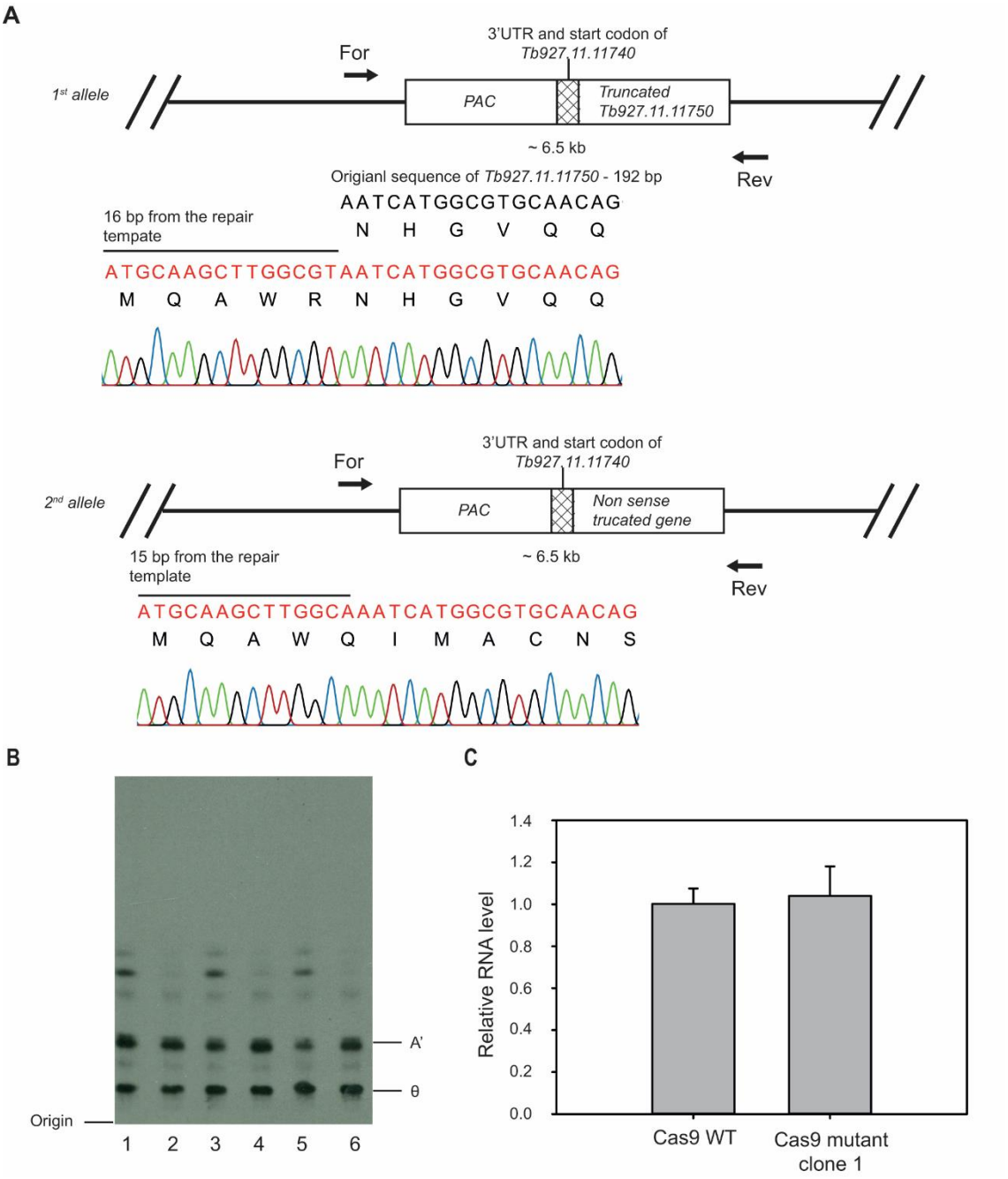

Figure 6.21 Characterisation of the genotypes and GPI anchor products of Cas9 mutant clone 1.

A: Sequences of the *Tb927.11.11740* and *Tb927.11.11750* gene loci following genetic rearrangements in the Cas9 mutant clone 1. Sequence analysis of 8 amplicon clones indicated that the two alleles were each modified in different ways: One allele contains the PAC selection marker followed by 500 bp of 3'UTR *Tb927.11.11740* targeting sequence, followed by 16 bp of the repair template sequence that places the *Tb927.11.11750* remnant sequence encoding (Δ1-64)Tb927.11.11750 in-frame with to potentially produce a MAQWR-(Δ1-64)Tb927.11.11750 fusion protein. The other allele has similar arrangement but with only 15 bp of repair template sequence giving rise to a nonsense truncated gene downstream of the potential Met start site. The forward and reverse primers used for amplifying the fragments for sequencing have been

annotated as “For” and “Rev”, respectively. B: Cell-free systems were labelled with GDP-[^{3}H]Man with or without a chase using cold GDP-mannose. The lipid extracts were analysed by HPTLC using the solvent described at Materials and Methods and analysed by fluorography. Glycolipid products A’ and θ are indicated. Lane 1 and 2: Cas9 WT samples; Lane 3 and 4: Cas9 mutant clone 1 samples; Lane 5 and 6: BSF WT samples. Lanes 1, 3 and 5 without cold GDP-Man chase, lanes 2, 4 and 6 with cold GDP-Man chase. C: Real time RTq-PCR showed similar levels of *Tb927.11.11750* mRNA in the Cas9 WT and Cas9 mutant clone 1 cell lines. Error bars indicate one standard deviation.

6.7 Attempts at the generation of *Tb927.11.11740* and *Tb927.11.11750* null and conditional null mutants through conventional homologous recombination

The CRISPR Cas9 mediated gene manipulation of *Tb927.11.11740* and *Tb927.11.11750* failed to generate a null mutant cell line for both genes simultaneously. Therefore, an alternative approach based on conventional homologous recombination was applied to try to make conditional null mutant cell lines. For the first step using this strategy, we attempted to replace the 1st allele of *Tb927.11.11740* and *Tb927.11.11750* with PCR amplicons containing the PAC gene flanked by 800 bp 5’UTR of *Tb927.11.11740* and 500 bp 3’UTR of *Tb927.11.11750* (same as the CRISPR Cas9 repair template used in Section 6.6) to generate a *ΔTb927.11.11740 & Tb927.11.11750::PAC*/*Tb927.11.11740 & Tb927.11.11750* single knock out cell line (sKO) (Fig 6.22A). To check for correct gene replacement, the PAC-resistant clones of were assessed by Southern blot probed sequentially with a PAC ORF probe and a GOI ORF probe. NcoI was used to digest the genomic DNA extracted from the cloned cell lines. The predicted locus for *Tb927.11.11740* (5725 bp) and *Tb927.11.11750* (3302 bp) and for the expected PAC-containing gene replacement product (4681 bp) are indicated in (Fig 6.22B). The Southern blot probed with the PAC probe (Fig 6.22C) showed the same single band detection in three mutant clones (lane 2-4) and no detection in the BSF WT sample (lane 1). However, the size of this detected band in the three mutant clones was not at the expected position, which was 4681 bp. Instead, a much larger DNA fragment was observed (>8.6kb). The same size band was observed in the three mutant clones when the same blot was subsequently stripped and probed with the GOI probe, along with bands consistent with the presence of endogenous copies of *Tb927.11.11740* and *Tb927.11.11750* (Fig 6.22D). The BSF WT sample showed the same expected bands for *Tb927.11.11740* and *Tb927.11.11750*. However, comparison

of the band intensities suggests that the three mutant clones may contain only one copy of *Tb927.11.11740* in its original locus but both copies of *Tb927.11.11750* in their original loci. It is clear from these data that simple gene replacement by homologous recombination has nit occurs in any of the three mutant clones. One possible explanation for the unusual genotypes generated is that the PAC resistance cassette has inserted into the locus to generate a tandem PAC *Tb927.11.11740* sequence upstream of the central NcoI site and displacing the 5' NcoI site. Whatever the explanation, this unexpected genetic rearrangement failed to support further planned steps for generating the *Tb927.11.11740* and *Tb927.11.11750* null and/or conditional null mutant cell lines (i.e., introducing an ectopic copy expressing *Tb927.11.11740* and *Tb927.11.11750* and replacing the second allele of endogenous locus). Therefore, we did not pursue these cell lines to perform more genetic modifications.

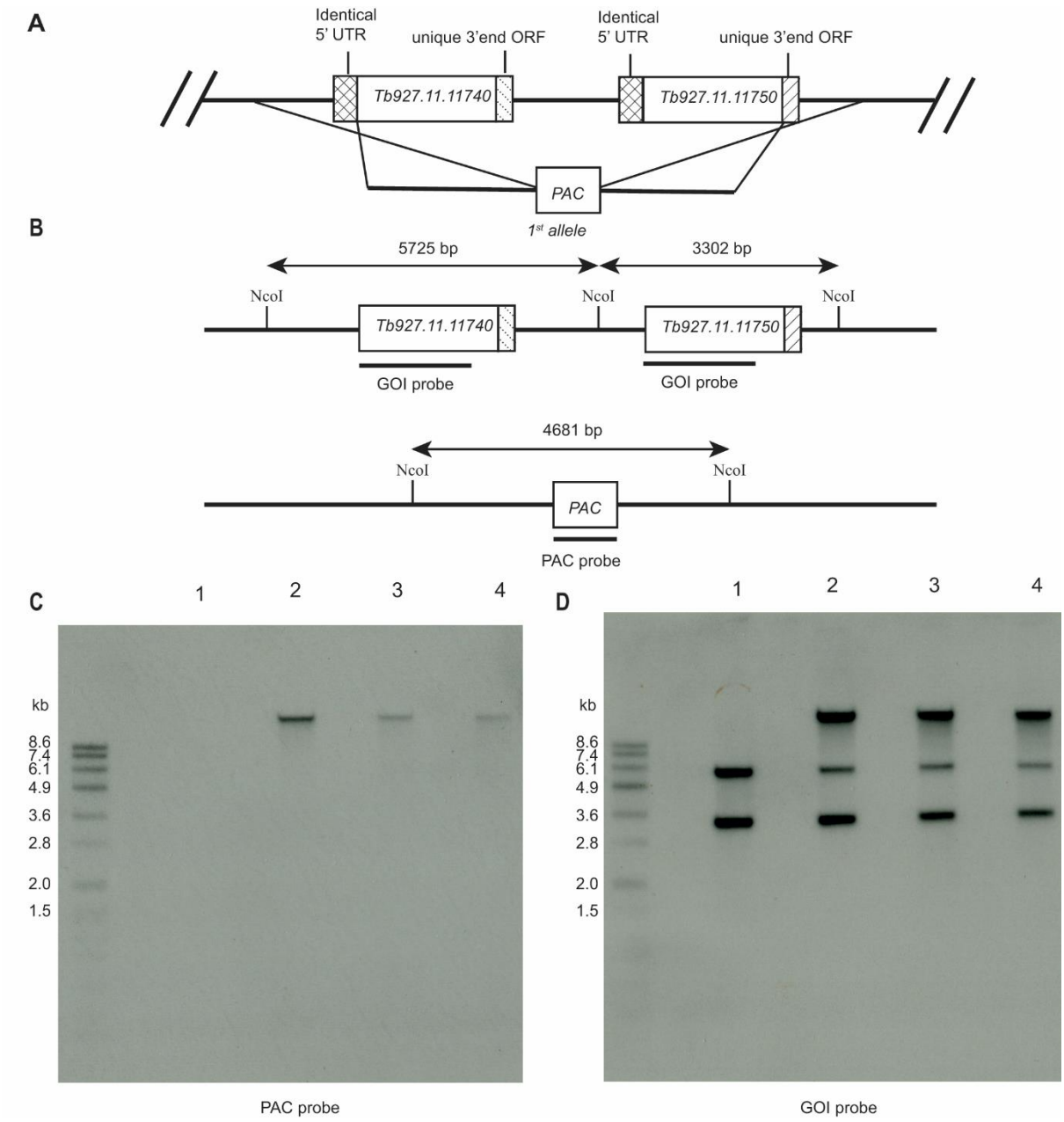

Figure 6.22 Attempt of generating *Tb927.11.11740* and *Tb927.11.11750* sKO cell lines.

A: To create the BSF single knock out for *Tb927.11.11740* and *Tb927.11.11750*, the first allele was targeted for replacement by PAC. B: Schematic representation of NcoI digestion sites and expected fragments to be detected by GOI and PAC probes for the *Tb927.11.11740* and *Tb927.11.11750* locus. The probes used for detection are indicated below the scheme of the gene locus, the expected size of bands detected by the probes are indicated above the scheme of the gene locus. C and D: Southern blot of BSF *T. brucei* gDNA (5 µg/ lane) of BSF WT (lane 1) and putative BSF *Tb927.11.11740* and *Tb927.11.11750* single knockout clones 1 to 3 (lanes 2-4) digested with NcoI. PAC probe (C) and GOI ORF (corresponding to 1000 bp identical region of *Tb927.11.11740* and *Tb927.11.11750*) (D) and were used for probing sequentially, with the blot stripped in between.

6.8 Attempts to generate a *Tb927.11.11750* conditional null mutant cell line through different conventional homologous recombination strategies

6.8.1 Attempt to generate *Tb927.11.11750* conditional null mutant cell line with an ectopic copy introduced with plew100_v1 *Tb927.11.11750*.

Since the CRISPR Cas9 mediated gene manipulation of *Tb927.11.11740* and *Tb927.11.11750* and our attempt to replace one allele of both genes simultaneously both resulted in abnormal gene rearrangements, we decided to simplify our gene editing strategy to focus on generating a *Tb927.11.11750* conditional null mutant. The planned strategy is described in (Fig 6.23). The Tb927.11.11750 ORF was amplified with primers OL13 and OL14 and inserted into the plew100 original version (plew100_v1) plasmid (Wirtz *et al.*, 1999). This allows the introduction of an ectopic copy of Tb927.11.11750 into the rRNA locus under the control of a tetracycline-inducible promoter using phleomycin selection. Following cloning and under induction with tetracycline, the first allele of *Tb927.11.11750* would be then replaced by homologous recombination with linear DNA containing hygromycin phosphotransferase (HYG) flanked by 500 bp 5'UTR and 3'UTR of *Tb927.11.11750*. After cloning and confirming the correct genotype of the conditional sKO cell line by Southern blotting, the second allele would be replaced with the puromycin resistance gene (PAC) by homologous recombination, generating the final BSF Δ*Tb927.11.11750::HYG*/Δ*Tb927.11.11750::PAC* conditional null mutant cell line (Fig 6.23A).

The PAC construct for knocking out (KO) *Tb927.11.11750* was synthesised by Genscript and the HYG construct was generated by Gibson assembly (Gibson *et al.*, 2009) as indicated in (Fig 6.23B). The DNA fragment that contained 5'UTR and 3'UTR of *Tb927.11.11750* was amplified from the synthesised pUC57-PAC plasmid with primers OL33 and OL34 in which the product would be lack of the PAC part. And the HYG gene was amplified using primers OL31 and OL32 from a HYG containing plasmid generated previously in lab. After Gibson assembly, the HYG construct for KO *Tb927.11.11750* was generated with HYG flanked by 500 bp 5'UTR and 3'UTRs of *Tb927.11.11750*. And the correct assembly of the plasmid was confirmed by DNA sequencing.

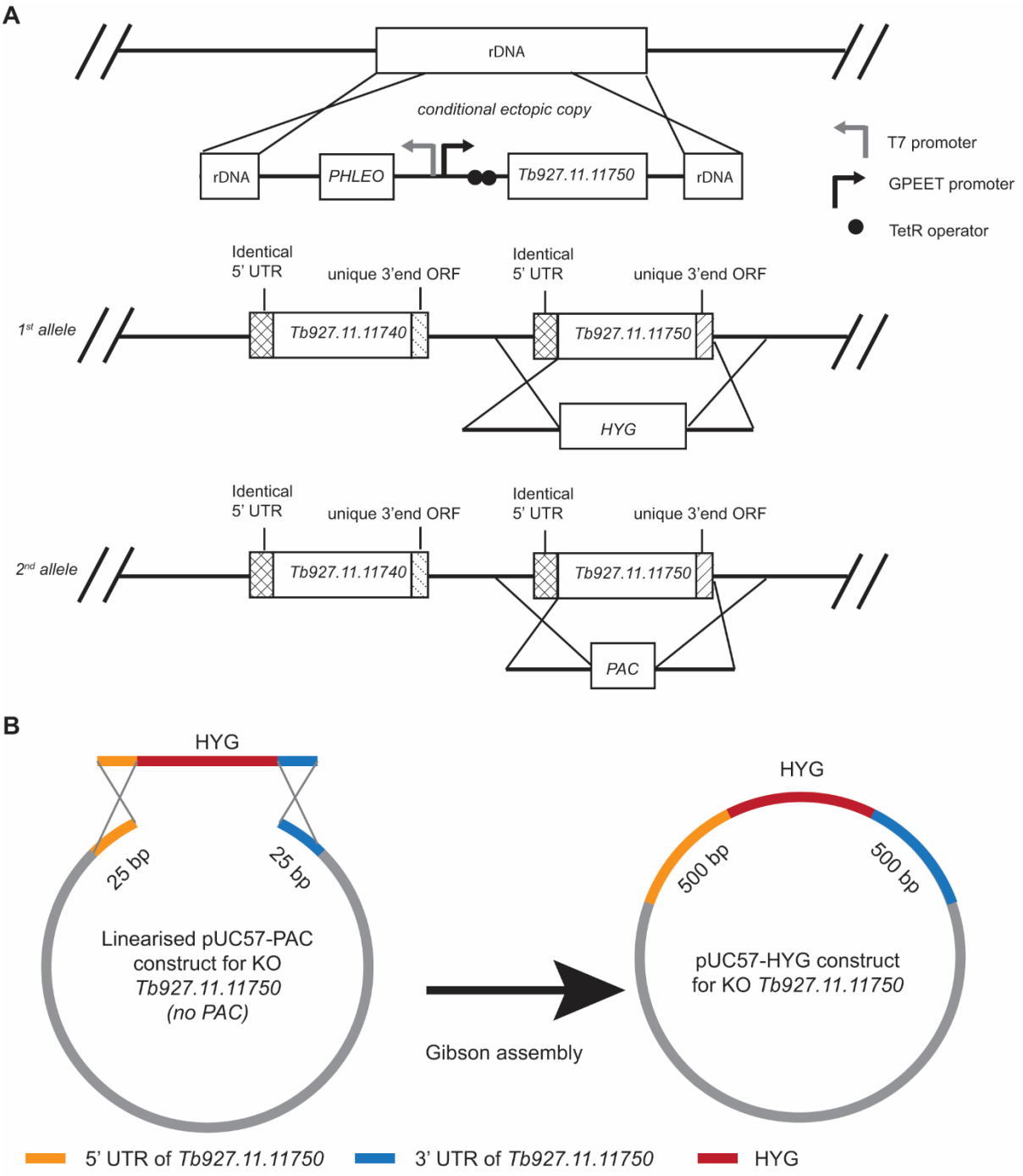

Figure 6.23 Planned strategy for the creation of the *Tb927.11.11750* conditional null mutant.

A: In order to create the BSF conditional null mutant for *Tb927.11.11750*, an ectopic copy of *Tb927.11.11750* that was tetracycline-inducible would be introduced into the ribosomal DNA locus, the first allele of *Tb927.11.11750* would be replaced by HYG followed by the replacement of second allele with PAC. B: Generation of HYG construct for knocking out *Tb927.11.11750* by Gibson assembly. Different DNA segments are annotated. HYG: hygromycin phosphotransferase.

To generate the construct of the ectopic copy of *Tb927.11.11750*, the open reading frame (ORF) of *Tb927.11.11750* was amplified from *T.brucei* gDNA using primers OL47 and OL48 This PCR product was further cloned to plew100_v1(Wirtz *et al.*, 1999) plasmid through HindIII and BamHI restriction enzyme sites and the correct insertion was verified through DNA sequencing. We chose the relatively low overexpression level vector plew100_v1 to try to mimic the low endogenous expression of Tb927.11.11750. To introduce an ectopic copy of *Tb927.11.11750*, *T. brucei* BSF WT cell line was transfected with NotI linearised plew100_v1 Tb927.11.11750 construct and after phleomycin selection and cloning, six clones were taken for genotype analysis by Southern blotting. The gDNA extracted from these six mutants and BSF WT were digested with NcoI and EcoRV to provide restriction fragments that represented endogenous *Tb927.11.11740* (5711 bp), endogenous *Tb927.11.11750* (3306bp) and the ectopic copy of *Tb927.11.11750* (2228 bp) (Fig 6.24 A). Despite imperfections of of the Southern blot, there were clearly two bands representing *Tb927.11.11740* and *Tb927.11.11750* detected in the BSF WT sample (Fig 6.24 B, lane 1) and an extra band for the ectopic copy of *Tb927.11.11750* was detected at the predicted position in all six overexpression clones (Fig 6.24B, lanes 2-7). This suggested that the ectopic copy of *Tb927.11.11750* was successfully introduced to *T.brucei* BSF WT in all clones. The overexpression level of clone 1 was then assessed through real time RT-qPCR in which the specific transcripts of *Tb927.11.11750* was ~2.5 fold higher in the plus Tet sample than in the minus Tet sample (Fig 6.24C) indicating a mild overexpression of *Tb927.11.11750* was achieved in this mutant cell line.

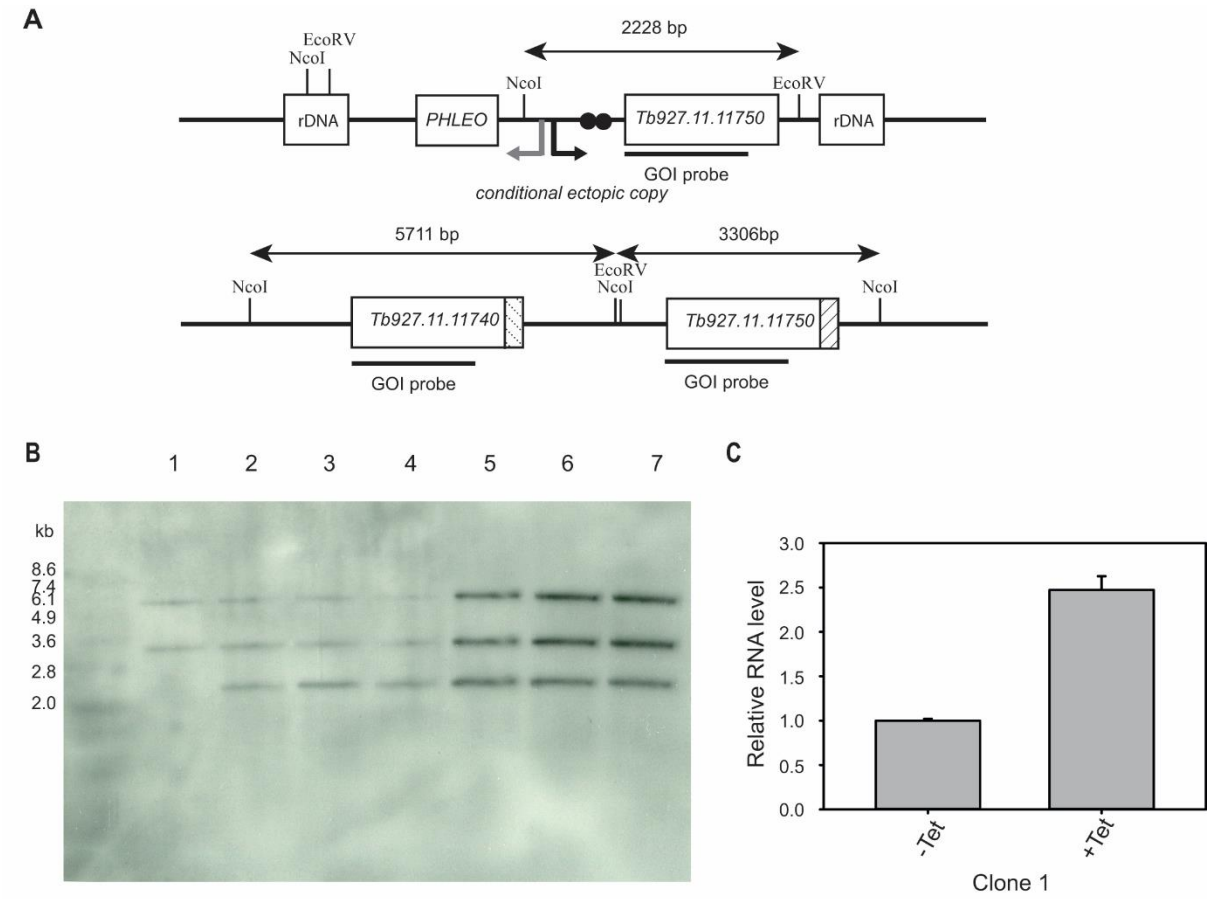

Figure 6.24 Introducing an ectopic copy of *Tb927.11.11750*.

A: Schematic representation of NcoI and EcoRV digestion sites and expected restriction fragments containing the ectopic copy of *Tb927.11.11750* and the endogenous copies of *Tb927.11.11740* and *Tb927.11.11750* as detected with a GOI probe. B*:* Southern blot of BSF *T.brucei* gDNA (5 µg/ lane) of BSF WT (lane 1) and BSF WT *Tb927.11.11750* overexpression mutant clones 1 to 6 (lanes 2-7) digested with NcoI and EcoRV. The GOI ORF (corresponding to 1000 bp identical region of *Tb927.11.11740* and *Tb927.11.11750*) was used for probing. C: Comparison of *Tb927.11.11750* specific transcript levels of clone 1 of *Tb927.11.11750* overexpression cell line without or with Tet induction examined by real time RTq-PCR. Error bars indicate done standard deviation.

After the induction of expression of the ectopic copy of *Tb927.11.11750*, the first endogenous allele of *Tb927.11.11750* was replaced with HYG flanked with 500 bp 5' and 3' UTRs of *Tb927.11.11750*. The generation of this HYG KO construct was described previously. The cells recovered from hygromycin selection were subcloned for the genotype analysis by Southern blot. For this analysis, endonuclease digestion by NcoI was performed on gDNAs extracted from *T.brucei* BSF cell lines, and the expected sizes of fragments to be detected with the GOI and HYG ORF probes are indicated in (Fig 6.25A). Since the GOI probe detects both *Tb927.11.11740* and

Tb927.11.11750 we would expect the intensity of the *Tb927.11.11750* band (at 3302 bp) to be half that of the *Tb927.11.11740* band (at 5725 bp). We could not predict the size of the ectopic copy of *Tb927.11.11750*, but by the comparison between the BSF WT sample and the plew100_v1 *Tb927.11.11750* overexpression sample we were able to determine that it resides within a 7.4-8.6 kb fragment (Fig 6.25B, lanes 1 and 2). We were also able to predict that when HYG is integrated to the correct locus, the HYG probe should detect two bands at 736 bp and 2124 bp (Fig 6.25A). For the Southern blot, gDNAs of nine clones recovered from HYG selection were prepared in parallel with BSF WT and plew100_v1 *Tb927.11.11750* overexpression cell lines for NcoI endonuclease reaction. The GOI probe detected the ectopic copy of *Tb927.11.11750* at 7.4-8.6 kb in all of the HYG selected clones (Fig 6.25B, lanes 3-11). Among these putative "sKO" mutants, there were four clones where the band corresponding to the *Tb927.11.11750* locus (3302 bp) was clearly half intensity of the band corresponding to *Tb927.11.11740* locus (5725 bp) (Fig 6.25B, lanes 3, 4, 8 and 9) corresponding to clones 1, 2, 7 and 7. This suggested that one allele of *Tb927.11.11750* was successfully replaced by HYG, the integration of which was confirmed in the HYG probe blot (see corresponding lanes with HYG probe in Fig 6.25B). However, in all cases there was another GOI band (annotated with asterisk) running below the endogenous *Tb927.11.11750* band in all nine clones, suggesting that replication and transposition of *Tb927.11.1150* or *Tb927.11.11740* also occurred during the replacement of *Tb927.11.11750* by HYG. As for the HYG probe blot, although the expected two bands (736 bp and 2124 bp) were detected, there was also another band (dagger symbol) detected suggesting that HYG was also incorporated into another locus in the genome. The two bands that were labelled with question mark were non-specific detections when using HYG probe as they were detected in the BSF WT sample as well.

While we do not fully understand the rearrangements that occurred in what should have been a simple gene replacement, possible explanations include: (i) Insufficient expression of the hygromycin resistance cassette by the *Tb927.11.11750* endogenous promotor, such that only clones that incorporated a second copy of HYG survived the selection. A similar problem was observed with the generation of a BSF GDP-fucose biosynthesis null mutant (Turnock *et al*., 2007). (ii) Haploid insufficiency of *Tb927.11.11750*, even though the activated ectopic copy apparently delivers more mRNA than two endogenous alleles (Fig 6.24C). (iii) There is something unusual about the *Tb927.11.11750* endogenous gene locus itself that makes gene replacement difficult.

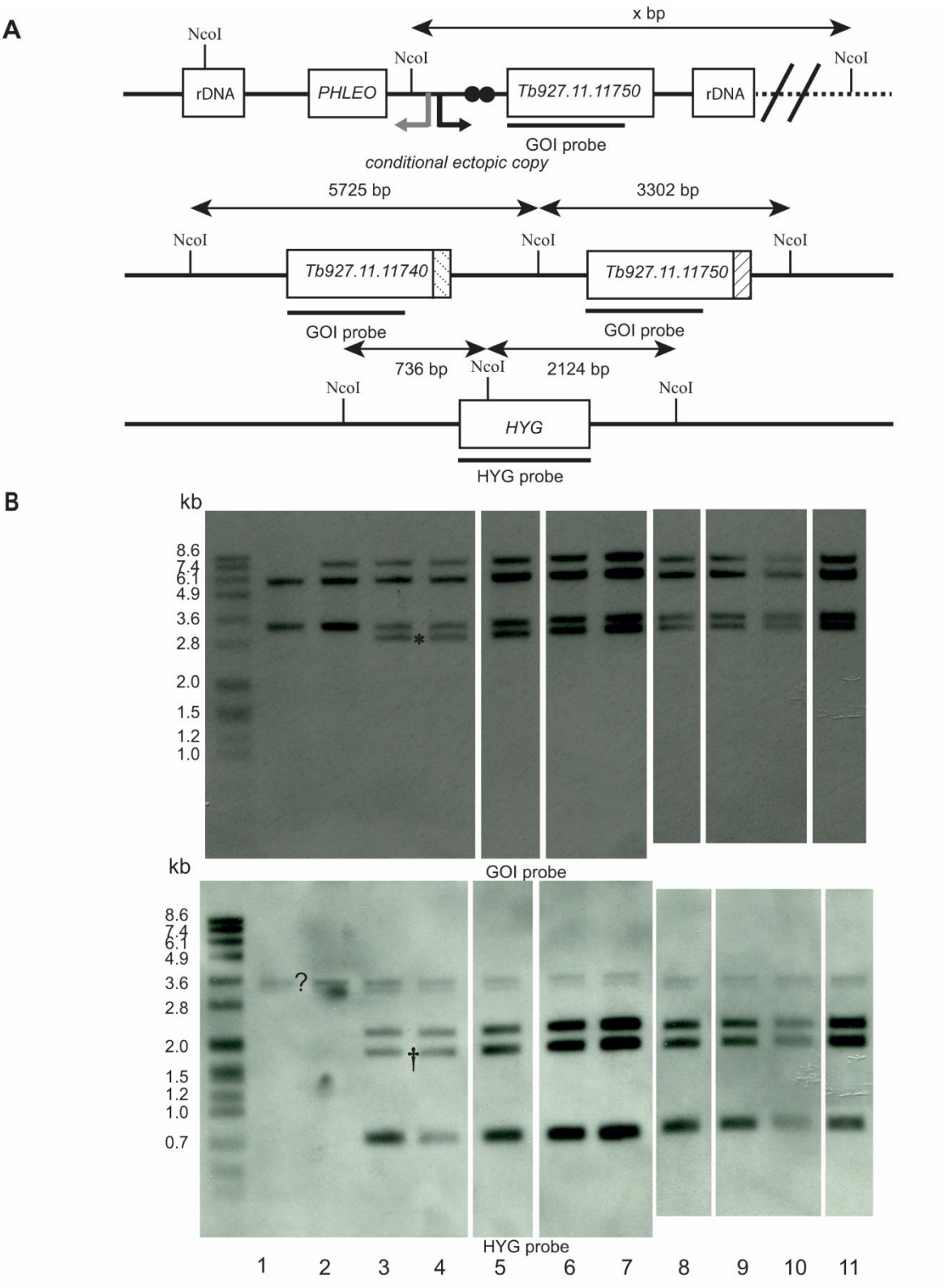

Figure 6.25 Attempt of generating *Tb927.11.11740* and *Tb927.11.11750* conditional sKO cell lines.

A: Schematic representation of NcoI digestion sites and expected probe detection (GOI and HYG probes) for the *Tb927.11.11740* and *Tb927.11.11750* locus. The probes used for detections was indicated below the scheme of gene locus, the expected size of bands detected by the probes were indicated above the scheme of gene locus. B: Southern blot of GDNA (5 µg/

lane) of BSF WT (lane 1) , BSF plew100_v1 *Tb927.11.11750* overexpression cell line (lane 2) and BSF conditional sKO mutants of *Tb927.11.11750* clones 1 to 9 (lanes 3-11) digested with NcoI. GOI ORF (corresponding to 1000 bp identical region of *Tb927.11.11740* and *Tb927.11.11750*) and HYG ORF probe were used for probing sequentially.

6.8.2 Attempt to generate a *Tb927.11.11750* conditional null mutant cell line using KO constructs with actin regulation elements

One of our hypotheses for why gene rearrangements occurred in (Fig.6.25 B, section 6.8.1) when using a conventional knock out method was that transcription of endogenous *Tb927.11.11750* locus in BSF WT might be insufficient to yield enough hygromycin resistance cassette to provide effective drug resistance during selection. To circumvent this situation, the PAC and HYG constructs for knocking out *Tb927.11.11750* were modified with the addition of 5'UTR and 3'UTR elements of the highly-expressed actin gene (the sequences were retrieved from plew100_v1 plasmid where they were used for regulating *PHLEO* resistance cassette expression. By doing this, we expected that higher levels of PAC and HYG transcripts and PAC and HYG protein would be achieved, hopefully preventing abnormal gene rearrangements when generating the conditional null mutant of *Tb927.11.11750* in BSF WT. A schematic representation of the creation of the conditional null mutant clones is shown in (Fig. 6.26A), note: these gene replacements were attempted in the same BSF plew100_v1 *Tb927.11.11750* overexpression cell line characterised in (Fig 6.24). After introducing PAC and HYG replacement constructs in two rounds of transfection and selection, gDNA from mutant clones was analysed by sequential Southern blotting using GOI, PAC and HYG ORF probes to check the genotype. Schematic detections for PAC and HYG ORF probes were shown in (Fig 6.26B) while the GOI ORF probe detection is indicated in (Fig 6.25A). Unfortunately, the blot with the GOI probe showed that the plew100_v1 *Tb927.11.11750* overexpression cell line had lost the ectopic copy after recovering from liquid nitrogen stabilate (Fig 6.25A, lane 2). The PAC conditional sKO clone 1 (Fig 6.26C, lane 3) showed loss of one allele of endogenous *Tb927.11.11750* as the intensity of the band corresponding to endogenous *Tb927.11.11750* (3302 bp) was half the intensity of the band corresponding to endogenous *Tb927.11.11740* (5725 bp). However, the same *Tb927.11.11750* band was also observed in the three putative conditional null mutant clones (Fig 6.26C, lanes 4-6) where the last *Tb927.11.11750*

allele should have been replaced if both the PAC and HYG replacement cassettes integrated to the correct loci. The blot probed with PAC ORF probe (Fig 6.26D) showed expected band (2858 bp) detected in the conditional sKO (lane 3) and putative conditional null mutant clones 1-3 (lanes 4-6) suggesting that the PAC resistance cassette had integrated to the correct locus. However, an extra band (asterisk labelled) was detected just below the expected band indicating that PAC resistance cassette inserted to another locus in the *T.brucei* genome. While the blot probed with HYG ORF probe showed HYG resistance cassette had integrated to the correct locus in the three mutant clones. Together, the Southern blot data showed that PAC and HYG resistance cassettes had replaced two alleles of the endogenous *Tb927.11.11750* even though that the PAC resistance cassette inserted to another locus, while there was still endogenous remaining *Tb927.11.11750* detected with the GOI ORF probe. This suggested that either there were more than two copies of *Tb927.11.11750* in the BSF WT genome to begin with, or (more likely) that a gene duplication and transposition of *Tb927.11.11750* occurred when replacing the two genes in the absence of an activated ectopic copy. In any case, given the complicated genotypes we elected to abandon this particular approach.

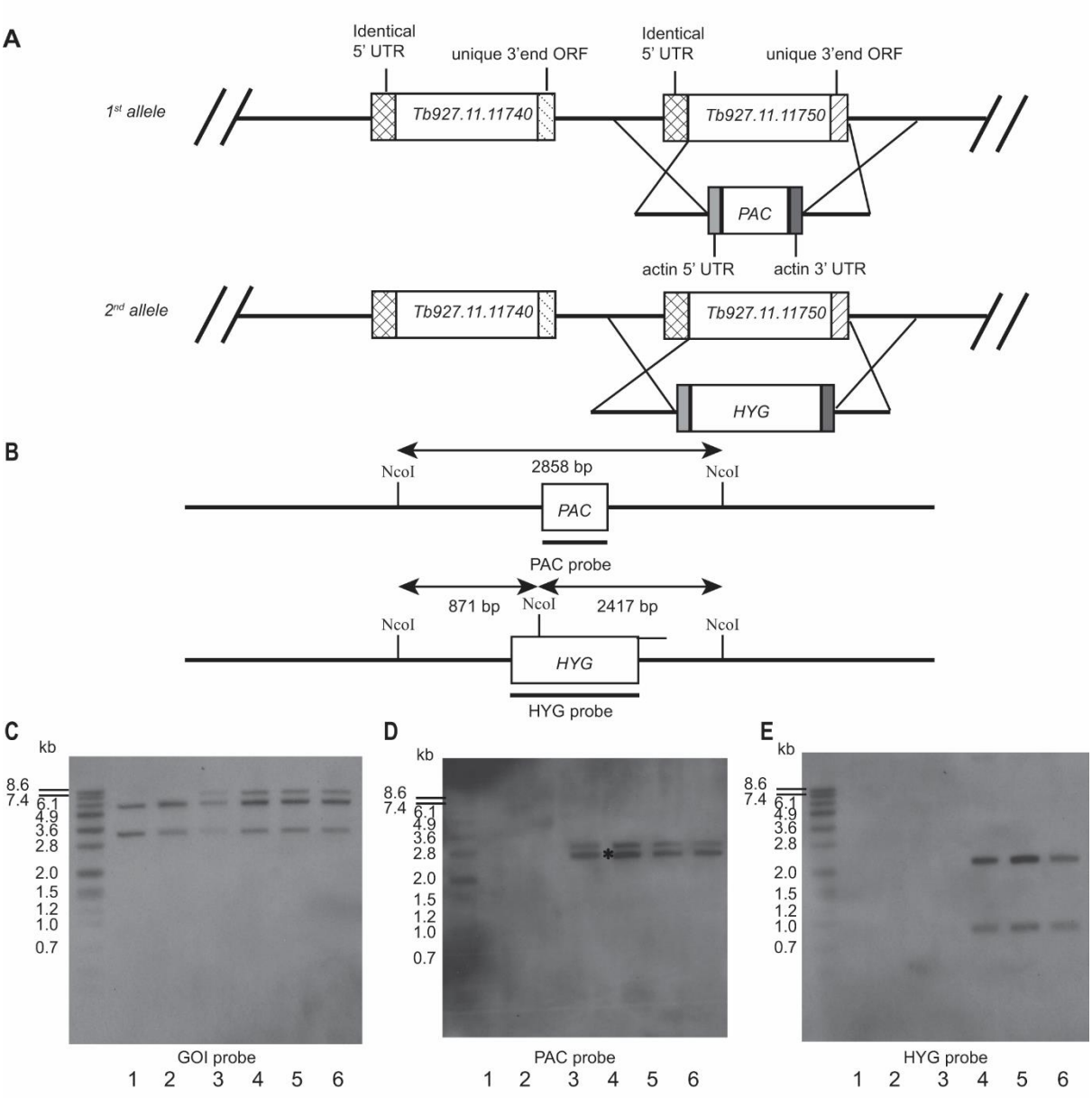

Figure 6.26 Schematic strategy for the creation of the *Tb927.11.11750* conditional null mutant.

A: The conditional null mutant for *Tb927.11.11750* was generated in the plew100_v1 overexpressing *Tb927.11.11750* BSF WT cell line described in Fig 6.24. Two alleles of endogenous *Tb927.11.11750* were replaced with PAC and HYG resistance cassette sequentially where the PAC and HYG were flanked with 5' UTR and 3' UTR of actin gene for high transcription of the resistance cassettes together with the 5'UTR and 3'UTR of *Tb927.11.11750* for homologous recombination. B: Schematic representation of NcoI digestion sites and expected probe detection (PAC and HYG probes) for the *Tb927.11.11750* locus. The probes used for detection are indicated below the scheme of the gene locus, the expected size of bands detected by the probes are indicated above the scheme of gene locus. C, D and E: Southern blots of gDNA (5 µg/ lane) extracted from BSF WT (lane 1), BSF plew100_v1 *Tb927.11.11750* overexpression cell line (lane 2), putative BSF conditional sKO clone 1 (lane 3) and putative

BSF conditional null mutants clones 1-3 (lanes 4-6). GOI ORF (corresponding to 1000 bp identical region of Tb927.11.11740 and Tb927.11.11750), PAC and HYG ORF probes were used for probing sequentially.

6.8.3 Attempt to generate a *Tb927.11.11750* conditional null mutant cell line with ectopic copy introduced using plew100_v5 *Tb927.11.11750*.

Our attempt to generate a *Tb927.11.11750* conditional null mutant using a plew100_v1 *Tb927.11.11750* ectopic copy had failed, and we wonder if it was the overexpression level of this ectopic copy *Tb927.11.11750* using plew100_v1 vector was not sufficient to compensate for the of the endogenous *Tb927.11.11750* copies. Therefore, the BSF WT was transformed with an ectopic copy of *Tb927.11.11750* using a plew100_v5 vector which generally results in about 10-fold higher Tet induced expression than plew100_1 constructs (Fig 6.28A). Primers OL15 and OL16 were used to amplify the of *Tb927.11.11750* ORF and OL17 and OL18 were used to amplify the linearised plew100_v5 vector. The gDNA of the clones recovered from the phleomycin selection were used for PCR analysis (Fig 6.28A) using the indicated primers. The five tested clones (Fig 6.28B, lanes 1-5) showed specific 2205 bp amplicons consistent with the successful integration of the ectopic copy into the *T. brucei* ribosomal DNA locus. The PCR amplification from the plew100_v5 *Tb927.11.11750* plasmid using the same pair of primers (Fig 6.28B, lane 6) was included as a positive control. The overexpression level of clone 1 mutant was then examined through real time RT-qPCR with or without 24 h Tet induction using primers OL37 and 38. Transcripts of *Tb927.11.11750* were ~50 fold higher in the plus Tet sample than in the minus Tet sample (Fig 6.28C), suggesting a high overexpression *Tb927.11.11750* mutant cell line was generated as expected. Then, the endogenous *Tb927.11.11750* were replaced by transfections of HYG and PAC integration constructs as indicated in the scheme of (Fig 6.23A). After two rounds drug selection and subcloning, gDNA from mutant clones was analysed by Southern blotting using the GOI ORF probe to check the genotype. Schematic detection for GOI ORF probe was indicated is (Fig 6.25A). The blot with the GOI probe showed, disappointingly, that although phleomycin resistance had been retained, the ectopic copy of *Tb927.11.11750* had been lost during the transformations (Fig 6.28D, lanes 1 and 2). The intensity of the band representing *Tb927.11.11750* was half the intensity of the band representing *Tb927.11.11740*, suggesting that HYG had integrated to one

allele of *Tb927.11.11750* locus (Fig 6.28D, lane 3). However, there was an extra band detected in what should have been the conditional single knock out (sKO) cell line, indicating a duplication and transposition of the *Tb927.11.11750* gene had occurred in this mutant cell line. The pattern was found in all the tested conditional "null mutant" clones (Fig 6.28D, lanes 4-7). These data show that we failed to obtain the expected conditional null mutant for *Tb927.11.11750* even when starting with a highly overexpressing ectopic copy of *Tb927.11.11750.* Indeed, it is possible that *Tb927.11.11750* overexpression causes a growth defect such that cells harbouring ectopic *Tb927.11.11750* can be outgrown by those that lose the transgene while retaining the selectable marker.

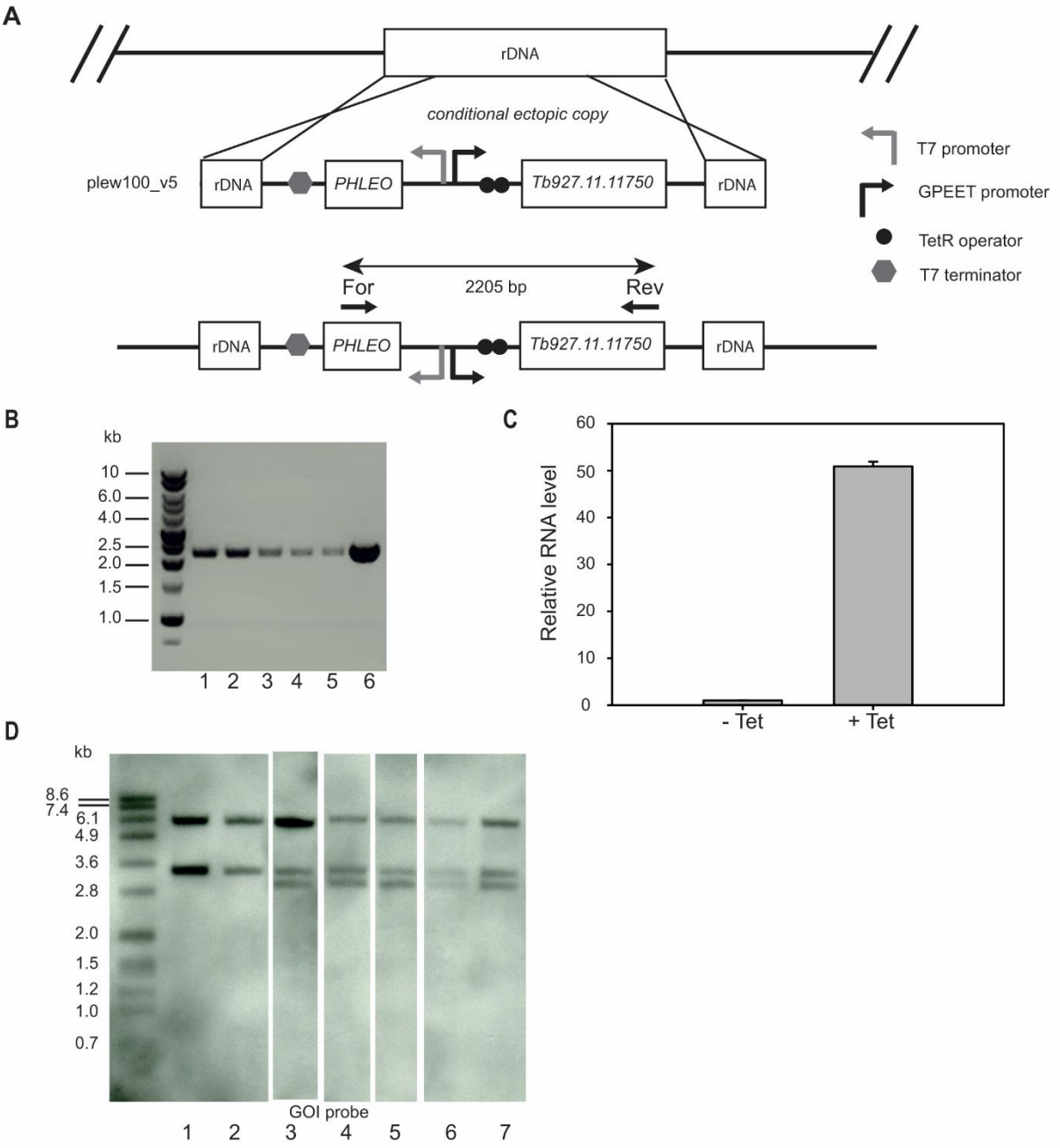

Figure 6.27 Attempt to generate a conditional null mutant of *Tb927.11.11750* using an ectopic copy delivered with plew100_v5 *Tb927.11.11750*.

A: Schematic representation of introducing an ectopic copy of plew100_v5 *Tb927.11.11750*. The insertion of the ectopic copy was analysed by PCR amplification of the gDNA of mutant clones giving 2205 bp product using specific primers (annotated with dark arrows): For, anneals to PHLEO region; Rev, anneals to *Tb927.11.11750* ORF region. B: PCR products from five clones (lanes 1-5) of plew100_v5 *Tb927.11.11750* overexpressing cell line were separated on agarose gel and analysed by electrophoresis. The PCR product using the plew100_v5 *Tb927.11.11750* plasmid as a template served as positive control. C: Comparison of *Tb927.11.11750* specific transcript levels of clone 1 of *Tb927.11.11750* overexpression cell line without or with 24 h Tet induction as determined by real time RTq-PCR. Error bars indicate one standard deviation. D: Southern blot of gDNA (5 µg/ lane) extracted from BSF WT (lane 1),

BSF plew100_v5 *Tb927.11.11750* overexpression clone 1 (lane 2), BSF putative conditional sKO clone 1 (lanes 3) and BSF putative conditional null mutants clones 1-4 (lanes 4-7) digested with NcoI. GOI ORF (corresponding to 1000 bp identical region of Tb927.11.11740 and Tb927.11.11750) probe was used for probing.

6.9 GPI PLA2 enzymatic assay test for Tb927.11.11740 and Tb927.11.11750

All of our attempts to generate a conditional null mutant for *Tb927.11.11740* and/or *Tb927.11.11750* had failed, therefore we could not apply biological function studies on Tb927.11.11740 and Tb927.11.11750 by comparing BSF WT and the *Tb927.11.11740* and *Tb927.11.11750* null mutant cell lines and their cell-free systems.

Our series of pull-down experiments coupled with proteomics analyses have elucidated that TbdeAc2 forms a complex with Tb927.11.11750, but not with Tb927.11.11740. However, due to the facts that Tb927.11.11740 and Tb927.11.11750 share 90% amino acids sequence identity, and since both share a lipase motif (GXSXG), we postulate that Tb927.11.11740 harbours a similar protein function as Tb927.11.11750 but in a different subcellular environment. As GPI phospholipase A2 (PLA2) is the enzyme catalysing the next step after TbdeAc2, one of our hypotheses is Tb927.11.11750 and/or Tb 927.11.11740 play the role of GPI PLA2 in the GPI anchor biosynthesis pathway and/or in the GPI anchor endocytic recycling pathway (Buxbaum *et al.*, 1996). To test this hypothesis, on-bead digestions of radiolabelled glycolipid A' which is the substrate of GPI PAL2 were performed with Tb927.11.11740-3HA and Tb927.11.11750-12Myc pull down products respectively.

To achieve this, a *T. brucei* cell-free system was prepared to perform the radiolabelling of GPI glycolipid intermediates in the GPI anchor biosynthetic pathway, as described previously (Fig. 6.18A). Briefly, the cells were subjected to hypotonic lysis and the membrane fraction (cell-free system) containing the GPI biosynthetic machinery was recovered by centrifugation. Pulse-labelling of the cell-free system with GDP-[^{3}H]Man, with a subsequent chase with unlabelled excess GDP-Man, was performed and the labelled glycolipids were extracted and an aliquot resolved by high-performance thin layer chromatography (HPTLC) and visualised by auto-fluorography. The silica plate containing glycolipids A' and θ products (that had not been sprayed with enhancer for auto-fluorography) was scraped where glycolipids A' and θ were expected and the glycolipids were recovered by extraction with 1-butanol. As shown in (Fig 6.29A), pure glycolipid A' (lane 2) and θ (lane 1) were purified with the A' product slightly

contaminated by another radiolabelled component. On-bead digestion of the purified A' product was set up to test Tb927.11.11740 and Tb927.11.11750 for GPI PLA2 enzymatic activity. For this assay, 2×10^8 BSF WT and Tb927.11.11740-3HA expressing cells were lysed with 1% TX-100 and immunoprecipitation was performed with anti-HA magnetic beads. After washing for 3 times, radiolabelled A' products were added to the beads for enzymatic reaction. The lipids were then extracted from these reactions after the incubation. Glycolipid A' product incubated with only anti-HA magnetic beads (Fig 6.29B, lane1) was added as a negative control, while digestion with bee venom PLA2 served as a positive control. These samples were analysed by HPTLC and visualised by auto-fluorography. Radiolabelled GPI anchor glycolipid mix was run in parallel to indicate the positions of glycolipids A' and θ. As shown in (Fig 6.29B), bee venom PLA2 could digest most glycolipid A' to θ (lane 4). However, there was no significant θ product produced in the Tb927.11.11740-3HA digestion sample. Only trace amounts of θ product was found in that sample and in all negative controls. We conclude from this experiment that Tb927.11.11740 does not have detectable GPI PLA2 enzymatic activity in this assay.

The same enzymatic assay was performed for Tb927.11.11750-12Myc pull down products that were immunoprecipitated with MycTrap™ beads. As for the Tb927.11.11740-3HA protein assay, the positive control bee venom PLA2 could digest most glycolipid A' to θ (Fig 6.29C, lane 3). However, the Tb927.11.11750-12Myc (lane 2) pull down products did not produce any obvious θ product from A', with only trace amounts of θ product present that was also detected in the MycTrap™ beads only digestion sample (lane 1). Together, our PLA2 enzymatic assay for Tb927.11.11740 and Tb927.11.11750 suggested that neither of these two proteins has GPI PLA2 enzymatic activity that is detectable under these assay conditions.

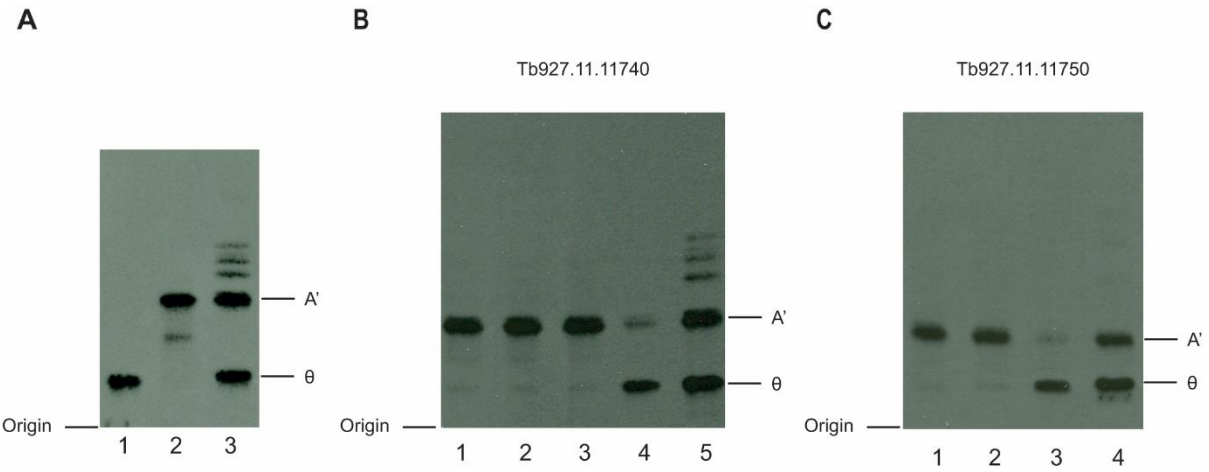

Figure 6.28 Tb927.11.11740-3HA and Tb927.11.11750-12Myc enzymatic assay.
A: HPTLC of purified radiolabelled glycolipids. Lane 1, purified glycolipid θ; lane 2, glycolipid A'; lane 3, radiolabelled GPI glycolipids mixture. B: Tb927.11.11740-3HA IP products on bead digestion test with glycolipid A'. Lane 1, glycolipid A' with anti-HA magnetic beads (negative control); lane 2, glycolipid A' incubated with BSF WT cell lysates IP products by anti-HA magnetic beads (negative control); lane 3, glycolipid A' incubated with Tb927.11.111740 cell lysates IP products by anti-HA beads; lane 4, glycolipid A' incubated with bee venom PLA2 (positive control); lane 5, radiolabelled GPI glycolipid mixture (standards). C: Tb927.11.11750-12Myc IP products on bead digestion test with glycolipid A'. Lane 1, glycolipid A' incubated with BSF WT cell lysates IP products by MycTrap™ beads (negative control). Lane 2, glycolipid A' incubated with Tb927.11.11150 cell lysates IP products by MycTrap™ beads. Lane 3, glycolipid A' incubated with bee venom PLA2 (positive control). Lane 4, radiolabelled GPI glycolipid mixture (standards).

7 Discussion

7.1 No evidence for direct protein interactions between TbGPI12 and TbMTI

The hypothesis that TbGPI12 might have associations with TbMTI came from studies on the acceptor specificities of GPI pathway enzymes. These studies used an assay based on the BSF trypanosome cell-free system described in (Section 5.3). The system was modified to include N-ethylmaleimide (NEM), which inhibits the endogenous GPI GnT of the pathway (Milne *et al*., 1992) So that it could be used for testing the abilities of exogenously added synthetic GlcNAc-PI and GlcN-PI, and analogues thereof, to prime the GPI pathway (Smith *et al.*, 1996). The read-out was based on the incorporation of [^{3}H]Man from GDP-[^{3}H]Man (via Dol-P-[^{3}H]Man) into the downstream GPI intermediates from Man_1GlcN-PI to glycolipid A'. In several such studies, GlcNAc-PI was always more efficient than GlcN-PI at priming the pathway. An example is shown in (Fig. 7.1).

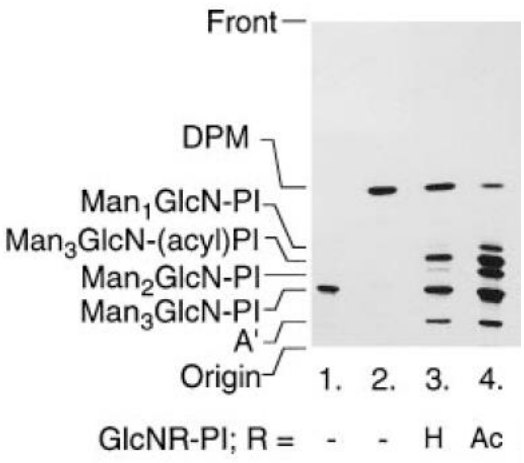

Fig 7.1 GlcNAc-PI is more efficient at priming the GPI pathway than GlcN-PI.
The HPTLC auto-fluorograph shown that the NEM-treated BSF cell-free system incubated with GDP-[^{3}H]Man makes only Dol-P-[^{3}H]Man (DPM) (lane 2) whereas the addition of synthetic GlcN-PI (lane 3) and GlcNAc-PI (lane 4) primes the production of downstream GPI intermediates, as indicated. Note the greater priming with GlcNAc-PI compared to GlcN-PI. These data are abstracted from Figure 2 of (Sharma *et al.*, 1997).

This phenomenon of the earlier intermediate (GlcNAc-PI) accessing TbMTI and the rest of the pathway more efficiently than its immediate substrate (GlcN-PI) suggested a

degree of substrate channelling between TbMTI and the preceding enzyme, the GlcNAc-PI de-N-acetylase (TbGPI12). Thus, we hypothesised that the two enzymes might form a protein complex. However, the immunoprecipitation experiments of TbGPI12-Myc coupled with either SILAC or cross-linking following by proteomics performed in this thesis failed to identify TbMTI as a partner protein for TbGPI12 (Section 4). Nor did a preliminary experiment described briefly in (Appendix A).

We also tried to perform the reverse pull-down by *in situ* tagging TbMTI with 3 x cMyc at the C-terminus to look for associated TbGPI12, this time in BSF cells (Appendix B). However, even starting with 1 e 9 cells we were barely able to detect the tagged protein in anti-Myc IPs (Appendix B, Fig S2) and so we did not follow this up.

In retrospect, since we now know that endogenous TbMTI is undetectable in our deep whole proteome dataset for BSF *T. brucei* (Ji, Tinti and Ferguson, 2021), it may be that we never had enough sensitivity to find TbMTI in overexpressed TbGPI12-Myc pulldowns. Thus, while we can provide no evidence for TbGPI12 – TbMTI association, we can still not rule it out.

Regardless, the native-PAGE anti-Myc Western blot of the TbGPI12-Myc IP product (Fig 4.4A) does suggest that TbGPI12 exists in a protein complexes, running between the 66 and 480 kDa markers. TbGPI12-Myc has a predicted MW of about 29 kDa and these complexes could be homo- or hetero-oligomers. After data filtering with stringent criteria for the proteomics data, the SILAC proteomics method identified 13 co-IP proteins of TbGPI12 (Table 4.1), while the cross-linking proteomics method identified 15 co-IP proteins (Table 4.2). The only protein that was present in both proteomics datasets is the *Tb927.10.14510* gene product which encodes a 95 kDa putative root hair defective 3 GTP-binding protein (RHD3). While it is not obvious why such a protein might be a component of the GPI pathway, one might speculate that is could play a regulatory role, as many GTP-binding proteins do. Any conclusions must, however, await more detailed studies, starting with the reverse pull down experiment.

Although our attempts to verify interactions between TbGPI12 and TbMTI failed, the IP studies using TbGPI12-Myc protein did benefit our later studies on other enzymes, in the GPI anchor biosynthesis pathway through: (i) The optimisation of the IP and detection conditions for low abundance membrane proteins; (ii) The set up for the SILAC proteomics identification and analysis methodology; (iii) Adoption of cross-linking methodology to study membrane protein-protein interactions. These conditions could also be applied to studying other membrane protein associations.

7.2 New components of GPI GnT in *T. brucei*

GPI anchor biosynthesis in all organisms is initiated with the transfer of GlcNAc from UDP-GlcNAc to phosphatidylinositol (PI) catalysed by the GPI GnT complex. The first component identified in this complex was PIGA in mammalian cells (Miyata *et al.*, 1993). The concept used in this paper was through a complementation experiment carried out in a GPI-deficient cell line JY5 with an Epstein-Barr virus vector-based cDNA library. Later, through affinity immunoprecipitation of PIGA, the PIGH subunit was found (Watanabe *et al.*, 1996). In this paper, the authors also proved that PIGA was the catalytic subunit and played its function on the cytoplasmic side of the ER lumen. Although the role of PIGH is not clear, it is essential for the activity of GPI GnT. The homologue of PIGC, GPI2, was first characterised in yeast, identified in a protein complex formed with GPI1 and GPI3 (Leidich *et al.*, 1995). After this, PIGC was also observed in mammalian cells (Inoue *et al.*, 1996). The yeast *GPI2* is an essential gene and PIGC is essential for GPI biosynthesis in mammalian cells. However, a recent study on the *T. brucei* homologue, *TbGPI2*, has shown that a null mutant of *TbGPI2* only exhibits a reduction in GPI GnT activity (Jenni *et al.*, 2021). *PIGQ* was found to be the mammalian homologue of yeast *GPI1* and proved to be a non-essential gene for GPI biosynthesis (Watanabe *et al.*, 1998). Another two components in this complex, PIGP and DPM2, were co-immunoprecipitated with PIGA in a two-step affinity purification. PIGP was shown to be an essential component for GPI GnT activity, while DPM is involved in the regulation of GPI GnT. In yeast, an *ERI1* gene product is another component for GPI GnT, whose activity is negatively regulated by Ras2 protein in the ER (Sobering *et al.*, 2004). The mammalian homologue of ERI1, PIGY, was also identified later (Murakami *et al.*, 2005). These studies have established that in mammalian cells, the GPI GlnT has at least seven components: PIGA, PIGH, PIGC, PIGQ, PIGP, PIGY and DPM2. In yeast, the GPI GnT complex contains all the corresponding homologues of these mammalian subunits except DPM2.

Although the studies on the GPI GnT complex are well established in mammalian and yeast cells, surprisingly, there was no experimental analysis for this complex in *T. brucei* before our study. With the established IP coupled with proteomics identification method described for TbGPI12 protein, we, therefore, performed IP for the epitope-tagged TbGPI3 protein, which is the catalytic subunit in this complex, followed by proteomics identification to explore the composition of the GPI GlnT complex in *T. brucei* and also to investigate whether any new components are specific for *T. brucei*.

The proteomics data in (Section 5) suggest that: (i) The *T. brucei* UDP-GlcNAc: PI α1–6 GlcNActransferase complex (GPI GnT) contains the expected subunits TbGPI3, TbGPI15, TbGPI19, TbGPI2, TbGPI1 and TbERI1. (ii) Like yeast, but unlike mammalian cells, DPM components are not subunits of the parasite complex. (iii) An Arv1-like protein (TbArv1) is a part of the parasite complex. (iv) A putative E2-ligase UbCE may be a part of the parasite complex.

The limitations of our study are that it does not attempt to assess stoichiometry, topology or inter-subunit associations of the *T. brucei* UDP-GlcNAc: PI α1–6 GlcNAc-transferase complex components and that it lacks an orthogonal confirmation of the presence of TbArv1 and TbUbCE in the complex. Regarding to the inter-subunit associations, the connections of the subunits for the mammalian cells has been reviewed in (Kinoshita, 2014). The study on TbGPI2 (Jenni *et al.*, 2021) in *T. brucei* has demonstrated that the absence of TbGPI2 causes the loss of most of TbGPI1 from the complex, indicating the important scaffolding role of TbGPI2 in the architecture of the GPI GnT complex in *T. brucei*.

With respect to the lack of an orthogonal confirmation of the presence of TbArv1 and TbUbCE in the complex, we rely, instead, on the extremely high and reproducible enrichment of these two components in the TbGPI3-3Myc pull-downs: From undetectable (>7149th) and 2538th in the total cell proteome for TbArv1 and TbUbCE, respectively, to 6th and 7th in the TbGPI3-3Myc pull-down proteome. TbArv1 is predicted to contain four transmembrane domains and an Arv1 domain (PF04161). Previous studies in yeast have indicated that Arv1p, although non-essential for growth and therefore GPI-anchoring of proteins at 25˚C (McDonough *et al.*, 2002) (Georgiev *et al.*, 2013), is required for the efficient synthesis of Man_1GlcN-(acyl)PI (Kajiwara *et al.*, 2008). It has been postulated to be a GPI flippase (Kajiwara *et al.*, 2008)(Okai *et al.*, 2020) helping deliver GlcN-(acyl)PI, which is made on the cytoplasmic face of the ER, to the active site of mannosyl-transferase I (MT I) on the luminal face of the ER. The complementation of yeast Arv1 mutants by the human Arv1(Ikeda *et al.*, 2016), and recent findings that human Arv1 mutations lead to deficiencies in GPI anchoring (Davids *et al.*, 2020) (Segel *et al.*, 2020), strongly suggest a related role in mammalian cells. However, whether it is a component of the mammalian and yeast GPI GnT complexes, or indeed of possible GlcNAc-PI de-N-acetylase or GPI flippase complexes, is unclear. It is possible that TbArv1 plays an analogous role to that proposed for Arv1 in yeast and mammalian cells in the *T. brucei* GPI pathway. However, since (unlike yeast and mammalian cells) acylation of the PI moiety occurs strictly after the action of

MT I in *T. brucei* (Güther and Ferguson, 1995), TbArv1 would need to facilitate the delivery of GlcN-PI rather than GlcN-(acyl)PI to TbMT I in this parasite. Further, it is worth noting that unlike mammalian cells and yeast, which are thought to only translocate GlcN-(acyl)PI, *T. brucei* appears to flip-flop most of its GPI intermediates between the cytoplasmic and lumenal surfaces of the ER (Vidugiriene and Menon, 1994). Thus, it is possible that *T. brucei* Arv1 protein, given its location, may play some other, perhaps regulatory, role in the GPI GnT reaction. This might also be the case in mammalian and yeast cells.

Finally, a recent study in mammalian cells showed that GPI anchor biosynthesis is upregulated in endoplasmic-reticulum-associated protein degradation (ERAD) deficient and PIGS mutant cell lines, suggesting that the GPI anchor biosynthetic pathway is somehow linked to and regulated by the ERAD system (Wang *et al.*, 2020). Since ERAD involves E2-dependent ubiquitylation of misfolded proteins as they exit the ER, it is possible that TbUbCE associated, in part, with the GPI GnT complex might play some role in regulation of the *T. brucei* GPI pathway.

Lastly, while for GPI GnT the chemical cross-linking, stringent IP proteomics approach (section 5.5) was less productive than the digitonin-solubilisation, native complex IP proteomics approach, it did identify one highly speculative new component, namely Tb927.9.14280. This predicted 187 kDa cytosolic protein contains a concanavalin A-like lectin domain that might, like Con A itself (Vidugiriene and Menon, 1994), be able to bind Man_2GlcN-PI and downstream GPI intermediates which, uniquely in *T. brucei*, occur on both sides of the ER membrane though flip-flop. Therefore, a hypothesis that might be tested in the future is to see whether Tb927.9.14280 provides negative feedback on the GPI GnT complex if GPI intermediates accumulate in the cytoplasmic face of the ER membrane. For example, if VSG synthesis was inhibited by cycloheximide, would glycolipid A and C levels greatly increase in absence of Tb927.9.14280? Tb927.9.14280 also contains an armadillo (ARM) repeat superfamily domain, roles for which include vesicular transport and membrane tethering, either or both of which could speculatively be relevant for ER membrane-based GPI biosynthesis.

7.3 Tb927.11.11750 - a possible regulatory role in GPI anchor biosynthesis in *T. brucei*

As described in the introduction and result chapters, enzymes in the GPI anchor biosynthesis pathway tend to work in complexes in different organisms. In *T. brucei*, the GPI GnT was experimentally shown in this thesis to contain at least six subunits (Ji, *et al*., 2021) and the GPI transamidase contains at least five components (Lillico et al., 2003) (Nagamune et al., 2003) (Hong, Nagamune, Ohishi, et al., 2006), in which TTA1 and 2 are unique subunits to *T. brucei* in terms of sequence but are topologically similar to PIGS/GPI17 and PIGU/GAB1 in mammals and yeast (Nagamune et al., 2003). The ethanolamine phosphate transferase has also been shown to be two component complex of TbGPI13, the catalytic subunit, stabilised by TbGPI11 (Stokes *et al.*, 2014). TbdeAc2, together with TbdeAc1, catalyse GPI inositol-deacylation of GPI intermediates (Fig 2.1). The first GPI inositol-deacylase to be identified in *T. brucei* was TbdeAc (Tb927.10.4780) (Güther *et al*., 2001). It was activity-labelled with the GPI inositol deacylation inhibitor [^{3}H]-diisopropylfluorophosphate (DFP) and purified. From peptide sequencing of the purified radiolabeled protein, TbdeAc1 was cloned and sequenced, showing high similarity with mammalian acyloxyacyl hydrolase. The conditional null mutant of TbdeAc1 in BSF showed that it is not essential, but its removal altered GPI precursor processing, leading to the accumulation on non-protein linked galactosylated free GPI glycolipids (Güther, Prescott and Ferguson, 2003). Immunofluorescence microscopy of C-terminal HA tagged TbdeAc1 suggested it is an ER-resident protein, although the co-localisation pattern with ER marker protein Bip was not homogeneous (Güther *et al*., 2003). In the BSF protein ranking (Ji, *et al*., 2021) TbdeAc1 ranks 2693rd out of 7149 identified protein groups, indicating that, with the exception of the transamidase complex, TbdeAc1 is a relatively abundant protein for the GPI pathway. We currently think that TbdeAc1 may be involved in the deacylation of excess glycolipid C to generate glycolipid A for galactosylation at a site distinct from that of GPI addition to VSG and other proteins.

TbdeAc2 (Tb927.3.2610) was identified through BLAST search using its mammalian homologue PGAP1 (Hong, Nagamune, Morita, *et al.*, 2006). The 70% knock-down of TbdeAc2 mRNA mediated by RNAi in BSF *T. brucei* caused accumulation of glycolipids C’ and lyso-C’ and a decrease of the non-inositol acylated inositol glycolipids A’ and θ. The inhibition of inositol deacylation also resulted in decreased surface expression of VSG, which caused cessation of growth – suggesting essentiality

of TbdeAc2 for BSF *T. brucei*. The phenotype of the TbdeAc2 knock-down is consistent with this enzyme catalysing the inositol-deacylation steps throughout the pathway leading to glycolipids A and C (Fig 1.10) and so we decided to focus on TbdeAc2 in this study.

Despite the aforementioned details of TbdeAc2 function, there are no subcellular localisation or protein complex data for this protein. Nor does TbdeAc2 appear in the proteome-wide protein complex database reported by our laboratory (Crozier *et al.*, 2017), probably because it is of low abundance (5715th out of 7149 proteins (Ji, *et al*., 2021)) and because the (Crozier *et al.*, 2017) study was enriched for soluble protein complexes.

Here, we have used immunofluorescence microscopy to localise TbdeAc2 to the ER, using an internal 12Myc *in situ* tagging approach to preserve its predicted N-terminal signal peptide and putative C-terminal ER retention signal. Further, we have provided evidence for specific protein-protein interaction between TbdeAc2 and Tb927.11.11750, and *vice versa*, through a series of *in situ* tagging and affinity immunoprecipitations followed by proteomics identification. The experiments performed are summarised in (Table 7.1).

Table 7.1 List of proteomics methods applied for TbdeAc2 and its partner proteins

Proteomics method	**Identified partner protein**
Label-free proteomics of TbdeAc2-3Myc IP products	No reliable co-IP protein
Cross-linking coupled with IP of TbdeAc2-3Myc followed by label-free proteomics	Tb927.11.11740 and/or Tb927.11.11750
SILAC proteomics identification of TbdeAc2-3Myc 0.5% digitonin lysis condition IP products	Tb927.11.11740 and/or Tb927.11.11750
SILAC proteomics identification of TbdeAc2-3Myc 0.3% digitonin lysis condition IP products	No reliable co-IP protein
Label-free proteomics of Tb927.11.11740-3HA IP products	No reliable co-IP protein
Label-free proteomics of SP-12Myc-Tb927.11.11750 IP products	TbdeAc2

It is worth noting that we could not distinguish Tb927.11.11750 from its close homologue Tb927.11.11740 by proteomics but, whereas we could IP TbdeAc2 with tagged Tb927.11.11750, we could not with tagged Tb927.11.11740.

The specific association between TbdeAc2 and Tb927.11.11750 strongly supports the notion that Tb927.11.11750 plays some role in the GPI anchor biosynthesis pathway in *T. brucei*. This could be either as an enzyme of the pathway itself, or as a regulatory and/or chaperone protein for TbdeAc2. With respect to the former possibility, we considered that Tb927.11.11750 (and its Tb927.11740 homologue) might encode GPI PLA2 activity since TbdAc2 and GPI PLA2 are adjacent activities in the GPI pathway (Fig 1.10) and since Tb927.11.11740 and Tb927.11.11750 contain GXSXG lipase motifs. Further, two versions of GPI PLA2 activity might be expected given that fatty acid remodelling occurs in the ER (Masterson et al., 1990) and fatty acid proof-reading of GPI-anchored VSG occurs in the endocytic system (Buxbaum *et al.*, 1994) . However, *in vitro* enzymatic assay of endogenously expressed Tb927.11.11750 and Tb927.11.11740 failed to demonstrate any GPI PLA2 activity on radiolabelled glycolipid A' substrate (Section 6.9). While GPI PLA2 activity cannot be ruled out for assay condition and sensitivity reasons, our enthusiasm for the GPI PLA2 hypothesis is much diminished by this result.

To enable us to determine the role of Tb927.11.11750 in GPI anchor biosynthesis pathway, our strategy was to generate either knockdown or conditional knock out mutants for *Tb927.11.11750*. If successful, this would allow us to radiolabel live trypanosomes depleted of Tb927.11.11750 with [^{3}H]myristic acid, [^{3}H]glucosamine and [^{3}H]mannose in the presence of tunicamycin, and their cell free systems with UDP-[^{3}H]GlcNAc and/or GDP-[^{3}H]Man in the presence of tunicamycin and with and without fatty acid remodelling mix (which provides myristoyl-CoA to drive the remodelling of glycolipid A' through to glycolipids A and C). Such experiments, together with chemical and enzymatic characterisation of the GPI glycolipid products observable by HPTLC and fluorography (Guther at al., 1994), would then allow us to pinpoint any GPI pathway defects due to the depletion of Tb927.11.11750.

We are able to obtain inducible RNAi clones that simultaneously knocked down *Tb927.11.11740* and *Tb927.11.11750*. However, the best clones we could select gave 70% knockdown using the pRPA approach (Table 7.2) (Alsford *et al.*, 2005). Unfortunately, when we made cell-free systems from these cells ±Tet, and from the parental cell-line, we were unable to detect any differences in GPI biosynthesis, as judged by GDP-[^{3}H]Man pulse and pulse-chase labelling (Fig 6.18B). However, this result was inconclusive since knockout rather

than knockdown of enzyme activity is frequently needed to see a phenotype. To give one recent example, in genome-wide (RIT-seq) RNAi knockdown (Alsford *et al.*, 2011), the novel mitochondrial fucosyltransferease (TbFUT1) showed no phenotype, whereas this gene proved to be essential for both BSF and PCF parasites using conditional null mutants for *TbFUT1*.

Table 7.2 List of attempts for generation knockdown or conditional null mutant cell line for *Tb927.11.11740* and/or *Tb927.11.11750*

Attempts of generation knock down/cKO cell line	**Mutant cell lines generated**
p2T7 RNAi knock down	Insufficient knockdown (25-40%)
pRPA RNAi knock down	70% knockdown after 24h Tet induction
CRISPR-Cas9 mediated gene knock out	Knock out *Tb927.11.11740* and intergenic region. Truncated *Tb927.11.11750*.
Conditional knock out of Tb927.11.11740 and Tb927.11.11750 simultaneously	Duplication of Tb927.11.11740 and Tb927.11.11750 gene locus
plew100_v1 ectopic copy conditional null mutant (Resistance Marker integrated with Tb927.11.11750 UTRs)	Duplicate gene locus of Tb927.11.11740 or Tb927.11.11750 was detected in the conditional sKO cell line.
plew100_v1 cKO (Resistance Marker integrated with actin gene UTRs)	Detection of duplicated Tb927.11.11740 or Tb927.11.11750 after two rounds of KO attempts
plew100_v5 cKO (Resistance Marker integrated with Tb927.11.11750 UTRs)	Detection of duplicated Tb927.11.11740 or Tb927.11.11750 in the conditional sKO cell line.

As described in (Section 6.6), we attempted to delete both alleles of *Tb927.11.11740* and *Tb927.11.11750* by CRISPR-Cas9 mediated gene editing (Rico *et al.*, 2018) using an sgRNA sequence common to both genes and a repair template containing PAC resistance cassette flanked with 800 bp 5'UTR of *Tb927.11.11740* and 500 bp 3'UTR of *Tb927.11.11750*. However, we ended up with clones with complex genotypes. The original study on CRISPR-Cas9 gene editing in *T. brucei* shows that the editing and repair of the aquaporin 2 locus could produce chimeras with the immediately adjacent and highly-homologous and aquaglyceroporin 3 gene (Rico *et al.*, 2018). A similar scenario may have occurred in our *Tb927.11.11740* and *Tb927.11.11750* CRISPR-Cas9 edited clones. All three clones (Fig 6.20) had gene re-arrangements in the original locus. The dissection of the edited region of mutant clone 1 by DNA sequencing showed integration of the PAC resistance cassette, taking out

Tb927.11.11740 but leaving in one allele a truncated versions of *Tb927.11.11750* encoding a (Δ1-64)Tb927.11.11750 protein in-frame with a potential methionine start site provided by the repair template. In the other allele the truncated *Tb927.11.11750* sequence was out of frame with the potential methionine start site. A cell-free system was prepared for this mutant clone to phenotype the GPI anchor biosynthesis pathway. Disappointingly, there were no changes in the GPI anchor intermediates detected in the mutant clone (Fig 6.21). This might be explained if the truncated (Δ1-64)Tb927.11.11750 protein performs the same function as intact Tb927.11.11750. Despite the complexity of these results, they did at least inform us that the *Tb927.11.11740* gene is non-essential and suggest that *Tb927.11.11750* may be essential.

We then tried to make conditional null mutants of the *Tb927.11.11740* and *Tb927.11.11750* locus and specifically of the *Tb927.11.11750* locus (Sections 6.7 and 6.8) and (Table 7.2). In all cases we failed to generate Tet-inducible conditional null mutants. It seems that, in general, while we can replace one copy of *Tb927.11.11750* with a selectable marker, this occurs with simultaneous incorporation of the selectable marker into another (unknown) locus *and* with the duplication and transposition of a copy of *Tb927.11.11750* to another locus. Further, although we were able to introduce Tet-inducible ectopic copies of *Tb927.11.11750* alone or together with *Tb927.11.11740*, the cells had a propensity to lose these transgenes while retaining the *PHLEO* selectable marker, making subsequent gene replacement of the endogenous alleles impossible if we assume that *Tb927.11.11750* is essential. This occurred whether using low expression level plew100_v1 or higher expression plew100_v5 vector. It is possible that, given the aforementioned low transcription of the *Tb927.11.11750* locus, even modest overexpression of *Tb927.11.11750* is negatively selected.

Regardless of the precise reasons for the genetic rearrangements we observed while trying remove *Tb927.11.11750* by CRISPR Cas9 editing, and during attempts to make *Tb927.11.11740* / *Tb927.11.11750* and *Tb927.11.11750* conditional null mutants, we are inclined to believe that these rearrangements point to *Tb927.11.11750* being an essential gene. Unfortunately, our inability to generate cells sufficiently starved of Tb927.11.11750 to see a growth and biochemical phenotype means that we are unable to pin down its exact function. On the balance of probabilities, however, we suggest that Tb927.11.11750 serves as an essential subunit of the TbdeAc2 complex, possibly performing regulatory or stabilising functions for the catalytic subunit. Testing this hypothesis must await successful generation of a conditional null mutant. One approach

we might be try is the inducible di-Cre system, as recently reported for the GlcNAc-transferase *TbGT1*0 gene (Duncan et al., 2021).

7.4 General discussion

Studies on GPI anchor structure and biosynthesis began more than three decades ago; the first complete GPI anchor structure was solved in 1988 in *T. brucei* (Ferguson *et al.*, 1988). The importance of the GPI anchor was first implied in this organism as it is the membrane-attachment method for the variant surface glycoproteins (VSGs) which act as the defensive coat for this human disease-causing organism. The GPI structure has proven to be a universal protein attachment mechanism through almost all eukaryotes. The power of molecular biology in mammalian and yeast systems facilitated the identification and understanding of the enzymes of the GPI pathway. These enzymes, and the diseases caused by mutations of these enzymes, have been reviewed extensively (Kinoshita, 2020)(Kinoshita, 2014). Most of the enzymes in the *T. brucei* GPI pathway could be found via their mammalian or yeast homologous, as summarised in (Table 1.1). However, for the parasite enzymes, including in the steps that are specific for *T. brucei* such as fatty acid remodelling, new approaches are needed to find unique components. The finding of these specific new components will complete our understanding of the GPI anchor biosynthesis pathway in *T. brucei*.

Our studies described in this thesis have provided some alternative ways to look for unknown components of the GPI anchor biosynthesis pathway; taking account of the protein properties of the known enzymes and exploiting current proteomics identification methods. Our study on the protein interactions between TbGPI12 and TbMTI suggests that these two proteins do not associate with each other, correcting prior assumptions. Nevertheless, the optimisation steps for immunoprecipitation of TbGPI12-Myc helped to set up procedures for the pull-down of low abundance transmembrane proteins and protein complexes. And we identified another novel protein that might form a protein complex with TbGPI12 which is Tb927.10.14510 (RDH3). This protein might play a regulatory role in the GPI anchor biosynthesis pathway. And generation of conditional null mutant cell lines could be conducted to assess its protein role. When the standardised IP procedure was then applied to the *T. brucei* GPI GnT complex, all the expected subunits (apart from DPM2) and two more possible new components were found. With the additional cross-linking coupled proteomics method, the protein (Tb927.9.14280) that contains an ConA-like domain has

been identified. Epitope tagging of this protein could be performed to conduct reverse pull-down which will first confirm the complex it forms with the GPI GnT complex. *In vitro* on beads digestion would be then set up to with the IP products of this tagged protein to assess its enzymatic activity or regulation function including in the GPI anchor biosynthesis pathway. The co-IP study performed on TbdeAc2 has identified a new partner protein for it which is Tb927.11.11750. However, the functional studies performed on this protein failed to give clear conclusions. *In vitro* enzymatic assay strongly suggests that it does not display GPI PLA2 activity. All attempts to generate a conditional null mutant of *Tb927.11.11750* ended up with gene re-arrangements at the locus, making it impossible to assay the phenotype through loss of expression. BLASTp analysis for the orthologs of Tb927.11.11750 shows it does not have any homologues in other species beyond *Trypanosoma cruzi*. The BLAST search using the Tb927.11.11750 amino acid sequence as query found TcSYL_0181470. The *T. cruzi* predicted protein contains a putative signal peptide and a transmembrane domain, like Tb927.11.11750. Since gene manipulation of Tb927.11.11750 indicates that it is likely an essential gene for *T. brucei*, it is possible that TcSYL_0181470 may also be an essential gene for *T. cruzi*. The recent example of potent membrane protein-protein interaction inhibitors successfully curing animal models of dengue fever (Kingwell, 2021) provides some encouragement that the TbdeAc2 – Tb927.11.11750 interaction might also be a druggable target for African trypanosomiasis and Chagas' disease.

8 References

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bryant, S.H., Canese, K., Church, D.M., Clark, K., DiCuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Johnson, M., Kelly, C., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Krasnov, S., Kuznetsov, A., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Ostell, J., O'Sullivan, C., Panchenko, A., Phan, L., Preussm, D., Pruitt, K.D., Rubinstein, W., Sayers, E.W., Schneider, V., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Starchenko, G., Tatusova, T.A., Trawick, B., Vakatov, D., Wang, Y., Ward, M.W., Wilbur, W.J., Yaschenko, E. and Zbicz, K. (2012) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, 41(D1), pp. D8–D20.

Acosta-Serrano, A., Vassella, E., Liniger, M., Renggli, C.K., Brun, R., Roditi, I. and Englund, P.T. (2001) 'The surface coat of procyclic Trypanosoma brucei: Programmed expression and proteolytic cleavage of procyclin in the tsetse fly', *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), pp. 1513–1518.

Alibu, V.P., Storm, L., Haile, S., Clayton, C. and Horn, D. (2005) 'A doubly inducible system for RNA interference and rapid RNAi plasmid construction in Trypanosoma brucei', *Molecular and Biochemical Parasitology*, 139(1), pp. 75–82.

Alsford, S. and Horn, D. (2008) 'Single-locus targeting constructs for reliable regulated RNAi and transgene expression in Trypanosoma brucei', *Molecular and Biochemical Parasitology*, 161(1), pp. 76–79.

Alsford, S., Kawahara, T., Glover, L. and Horn, D. (2005) 'Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects', *Molecular and Biochemical Parasitology*, 144(2), pp. 142–148.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C. and Horn, D. (2011) 'High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome', *Genome Research*, 21(6), pp. 915–924.

Ashida, H., Hong, Y., Murakami, Y., Shishioh, N., Sugimoto, N., Kim, Y.U., Maeda, Y. and Kinoshita, T. (2005) 'Mammalian PIG-X and Yeast Pbn1p Are the Essential

Components of Glycosylphosphatidylinositol-Mannosyltransferase I', *Molecular Biology of the Cell*, 16(3), pp. 1439–1448.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J., Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington, C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoeckert, C.J., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G. and Wang, H. (2010) 'TriTrypDB: a functional genomic resource for the Trypanosomatidae', *Nucleic Acids Research*, 38(suppl_1), pp. D457--D462.

Autheman, D., Crosnier, C., Clare, S., Goulding, D.A., Brandt, C., Harcourt, K., Tolley, C., Galaway, F., Khushu, M., Ong, H., Romero-Ramirez, A., Duffy, C.W., Jackson, A.P. and Wright, G.J. (2021) 'An invariant Trypanosoma vivax vaccine antigen induces protective immunity', *Nature*, 595(7865), pp. 96–100.

Bangs, J.D., Uyetake, L., Brickman, M.J., Balber, A.E. and Boothroyd, J.C. (1993) 'Molecular cloning and cellular localization of a BiP homologue in trypanosoma brucei. Divergent ER retention signals in a lower eukaryote', *Journal of Cell Science*, 105(4), pp. 1101–1113.

Bartossek, T., Jones, N.G., Schäfer, C., Cvitković, M., Glogger, M., Mott, H.R., Kuper, J., Brennich, M., Carrington, M., Smith, A.S., Fenz, S., Kisker, C. and Engstler, M. (2017) 'Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat', *Nature Microbiology*, 2(11), pp. 1523–1532.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil,

D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S. and Zhang, J. (2021) 'UniProt: The universal protein knowledgebase in 2021', *Nucleic Acids Research*, 49(D1), pp. D480–D489.

Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J. and Madden, T.L. (2012) 'Domain enhanced lookup time accelerated BLAST', *Biology Direct*, 7, pp. 1–14.

Borges, A.R., Engstler, M. and Wolf, M. (2021) '18S rRNA gene sequence-structure phylogeny of the Trypanosomatida (Kinetoplastea, Euglenozoa) with special reference to Trypanosoma', *European Journal of Protistology*, 81, p. 125824.

Brands, A. and Ho, T.H.D. (2002) 'Function of a plant stress-induced gene, HVA22. Synthetic enhancement screen with its yeast homolog reveals its role in vesicular traffic', *Plant Physiology*, 130(3), pp. 1121–1131.

Burki, F., Roger, A.J., Brown, M.W. and Simpson, A.G.B. (2020) 'The New Tree of Eukaryotes', *Trends in Ecology & Evolution*, 35(1), pp. 43–55.

Buxbaum, L.U., Milne, K.G., Werbovetz, K.A. and Englund, P.T. (1996) 'Myristate exchange on the Trypanosoma brucei variant surface glycoprotein', *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp. 1178–1183.

Buxbaum, L.U., Raper, J., Opperdoes, F.R. and Englund, P.T. (1994) 'Myristate Exchange', *Journal of Biological Chemistry*, 269(48), pp. 30212–30220.

Casas-Sánchez, A., Perally, S., Ramaswamy, R., Haines, L.R., Rose, C., Yunta, C., Aguilera-Flores, M., Lehane, M.J., Almeida, I.C., Boulanger, M.J. and Acosta-Serrano, A. (2018) 'The crystal structure and localization of Trypanosoma brucei invariant surface glycoproteins suggest a more permissive VSG coat in the tsetse-transmitted metacyclic stage', *bioRxiv* [Preprint].

Chang, T., Milne, K.G., Güther, M.L.S., Smith, T.K. and Ferguson, M.A.J. (2002) 'Cloning of Trypanosoma brucei and Leishmania major genes encoding the GlcNAc-phosphatidylinositol De-N-acetylase of glycosylphosphatidylinositol biosynthesis that is essential to the African sleeping sickness parasite', *Journal of Biological Chemistry*,

277(51), pp. 50176–50182.

Chattopadhyay, A., Jones, N.G., Nietlispach, D., Nielsen, P.R., Voorheis, H.P., Mott, H.R. and Carrington, M. (2005) 'Structure of the C-terminal domain from Trypanosoma brucei variant surface glycoprotein MITat1.2', *Journal of Biological Chemistry*, 280(8), pp. 7228–7235.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) 'Multiple sequence alignment with the Clustal series of programs', *Nucleic Acids Research*, 31(13), pp. 3497–3500.

Clayton, C.E. and Mowatt, M.R. (1989) 'The procyclic acidic repetitive proteins of Trypanosoma brucei. Purification and post-translational modification', *Journal of Biological Chemistry*, 264(25), pp. 15088–15093.

Cross, G.A.M., Kim, H.S. and Wickstead, B. (2014) 'Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427', *Molecular and Biochemical Parasitology*, 195(1), pp. 59–73.

Crozier, T.W.M., Tinti, M., Larance, M., Lamond, A.I. and Ferguson, M.A.J. (2017) 'Prediction of protein complexes in trypanosoma brucei by protein correlation profiling mass spectrometry and machine learning', *Molecular and Cellular Proteomics*, 16(12), pp. 2254–2267.

Davids, M., Menezes, M., Guo, Y., Mclean, S.D., Hakonarson, H., Collins, F., Worgan, L., Billington, C.J., Maric, I., Okashah, R., Onyekweli, T., Adams, D.R., Christine, M. and Malicdan, V. (2020) 'Homozygous splice-variants in human ARV1 cause GPI-anchor synthesis deficiency', *Molecular Genetics and Metabolism*, 130(1), pp. 49–57.

Dean, S., Sunter, J.D. and Wheeler, R.J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource', *Trends in Parasitology*, 33(2), pp. 80–82.

Derewenda, Z.S. and Sharp, A.M. (1993) 'News from the interface: the molecular structures of triacyglyceride lipases', *Trends in Biochemical Sciences*, 18(1), pp. 20–25.

Duncan, S.M., Nagar, R., Damerow, M., Yashunsky, D. V., Buzzi, B., Nikolaev, A. V. and Ferguson, M.A.J. (2021) 'A Trypanosoma brucei β3 glycosyltransferase superfamily gene encodes a β1-6 GlcNAc-transferase mediating N-glycan and GPI anchor modification.', *Journal of Biological Chemistry*, p. 101153.

Eisenhaber, B., Bork, P. and Eisenhaber, F. (1999) 'Prediction of potential GPI-modification sites in proprotein sequences', *Journal of Molecular Biology*, 292(3), pp. 741–758.

Engstler, M., Reuter, G. and Schauer, R. (1993) 'The developmentally regulated trans-

sialidase from Trypanosoma brucei sialylates the procyclic acidic repetitive protein', *Molecular and Biochemical Parasitology*, 61(1), pp. 1–13.

Ferguson, Michael A.J., Hart Gerald W., K.T. (2017) 'Essentials of Glycobiology', in Ajit Varki, Executive Editor, Richard D Cummings, Jeffrey D Esko, Pamela Stanley, Gerald W Hart, Markus Aebi, Alan G Darvill, Taroh Kinoshita, Nicolle H Packer, James H Prestegard, Ronald L Schnaar, and P.H.S. (ed.). Cold Spring Harbor Laboratory Press.

Ferguson, M.A.J. (1999) 'The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research', *Journal of Cell Science*, 112(17), pp. 2799–2809.

Ferguson, M.A.J., Homans, S.W., Dwek, R.A. and Rademacher, T.W. (1988) 'Glycosyl-phosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane', *Science*, 239(4841), pp. 753–759.

Ferguson MAJ, Kinoshita T, H.G.W. (2009) 'Glycosylphosphatidylinositol Anchors', in et al. Varki A, Cummings RD, Esko JD (ed.) *Essentials of Glycobiology.* 2nd edn. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 137–150.

Field, M.C., Menon, A.K. and Cross, G.A. (1992) 'Developmental variation of glycosylphosphatidylinositol membrane anchors in Trypanosoma brucei. In vitro biosynthesis of intermediates in the construction of the GPI anchor of the major procyclic surface glycoprotein.', *Journal of Biological Chemistry*, 267(8), pp. 5324–5329.

Field, M.C., Menon, A.K. and Cross, G.A.M. (1991) 'A glycosylphosphatidylinositol protein anchor from procyclic stage Trypanosoma brucei : lipid structure and biosynthesis', 10(10), pp. 2731–2739.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Simarro, P.P., Zhao, W. and Argaw, D. (2020) 'Monitoring the elimination of human African trypanosomiasis at continental and country level: Update to 2018', *PLoS Neglected Tropical Diseases*, 14(5), pp. 1–18.

Freymann, D.M., Metcalf, P., Turner, M. and Wiley, D.C. (1984) '6 -Resolution X-ray structure of a variable surface glycoprotein from Trypanosoma brucei', *Nature*, 311(5982), pp. 167–169.

Fujita, M. and Kinoshita, T. (2012) 'Biochimica et Biophysica Acta GPI-anchor remodeling : Potential functions of GPI-anchors in intracellular traf fi cking and membrane dynamics ☆', *BBA - Molecular and Cell Biology of Lipids*, 1821(8), pp. 1050–1058.

Georgiev, A.G., Johansen, J., Ramanathan, V.D., Sere, Y.Y., Beh, C.T. and Menon, A.K. (2013) 'Arv1 regulates PM and ER membrane structure and homeostasis but is dispensable for intracellular sterol transport', *Traffic*, 14(8), pp. 912–921.

Gerber, L.D., Kodukula, K. and Udenfriend, S. (1992) 'Phosphatidylinositol glycan (PI-G) anchored membrane proteins. Amino acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide', *Journal of Biological Chemistry*, 267(17), pp. 12168–12173.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods*, 6(5), pp. 343–345.

Grünfelder, C.G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.D., Boshart, M. and Overath, P. (2002) 'Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels', *Traffic*, 3(8), pp. 547–559.

Gupta, R. and Brunak, S. (2002) 'Prediction of glycosylation across the human proteome and the correlation to protein function.', *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 322, pp. 310–322.

Güther, M.L. and Ferguson, M.A.J. (1995) 'The role of inositol acylation and inositol deacylation in GPI biosynthesis in Trypanosoma brucei.', *The EMBO Journal*, 14(13), pp. 3080–3093.

Guther, M.L.S. (2001) 'Purification, cloning and characterization of a GPI inositol deacylase from Trypanosoma brucei', *The EMBO Journal*, 20(17), pp. 4923–4934.

Güther, M.L.S., Lee, S., Tetley, L., Acosta-Serrano, A. and Ferguson, M.A.J. (2006) 'GPI-anchored Proteins and Free GPI Glycolipids of Procyclic Form Trypanosoma brucei Are Nonessential for Growth, Are Required for Colonization of the Tsetse Fly, and Are Not the Only Components of the Surface Coat', *Molecular Biology of the Cell*. Edited by J. York, 17(12), pp. 5265–5274.

Güther, M.L.S., Masterson, W.J. and Ferguson, M.A.J. (1994) 'The effects of phenylmethylsulfonyl fluoride on inositol-acylation and fatty acid remodeling in African trypanosomes', *Journal of Biological Chemistry*, 269(28), pp. 18694–18701.

Güther, M.L.S., Prescott, A.R. and Ferguson, M.A.J. (2003) 'Deletion of the GPIdeAc Gene Alters the Location and Fate of Glycosylphosphatidylinositol Precursors in Trypanosoma brucei †', *Biochemistry*, 42(49), pp. 14532–14540.

Güther, M.L.S., Treumann, A. and Ferguson, M.A.J. (1996) 'Molecular species analysis and quantification of the glycosylphosphatidylinositol intermediate glycolipid C from Trypanosoma brucei', *Molecular and Biochemical Parasitology*, 77(2), pp. 137–145.

Hermanson, G.T. (2013) ‘Homobifunctional Crosslinkers’, *Bioconjugate Techniques*, pp. 275–298.

Hillis, D.M. (1997) ‘Phylogenetic analysis’, *Current Biology*, 7(3), pp. R129–R131.

Hirata, T., Mishra, S.K., Nakamura, S., Saito, K., Motooka, D., Takada, Y., Kanzawa, N., Murakami, Y., Maeda, Y., Fujita, M., Yamaguchi, Y. and Kinoshita, T. (2018) ‘Identification of a Golgi GPI-N-acetylgalactosamine transferase with tandem transmembrane regions in the catalytic domain’, *Nature Communications*, 9(1).

Hirumi, H. and Hirumi, K. (1989) ‘Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers’, *The Journal of Parasitology*, 75(6), p. 985.

Hong, Y., Maeda, Y., Watanabe, R., Inoue, N., Ohishi, K. and Kinoshita, T. (2000) ‘Requirement of PIG-F and PIG-O for transferring phosphoethanolamine to the third mannose in glycosylphosphatidylinositol’, *Journal of Biological Chemistry*, 275(27), pp. 20911–20919.

Hong, Y., Nagamune, K., Morita, Y.S., Nakatani, F., Ashida, H., Maeda, Y. and Kinoshita, T. (2006) ‘Removal or maintenance of inositol-linked acyl chain in glycosylphosphatidylinositol is critical in trypanosome life cycle’, *Journal of Biological Chemistry*, 281(17), pp. 11595–11602.

Hong, Y., Nagamune, K., Ohishi, K., Morita, Y.S., Ashida, H., Maeda, Y. and Kinoshita, T. (2006) ‘TbGPI16 is an essential component of GPI transamidase in Trypanosoma brucei’, *FEBS Letters*, 580(2), pp. 603–606.

Horn, D. (2014) ‘Antigenic variation in African trypanosomes’, *Molecular and Biochemical Parasitology*, 195(2), pp. 123–129.

Ikeda, A., Kajiwara, K., Iwamoto, K., Makino, A., Kobayashi, T., Mizuta, K. and Funato, K. (2016) ‘Complementation analysis reveals a potential role of human ARV1 in GPI anchor biosynthesis’, *Yeast*, 33(2), pp. 37–42.

Inoue, N., Kinoshita, T., Orii, T. and Takeda, J. (1993) ‘Cloning of a human gene, PIG-F, a component of glycosyl- phosphatidylinositol anchor biosynthesis, by a novel expression cloning strategy’, *Journal of Biological Chemistry*, 268(10), pp. 6882–6885.

Inoue, N., Watanabe, R., Takeda, J. and Kinoshita, T. (1996) ‘PIG-C , One of the Three Human Genes Involved in the First Step of Glycosylphosphatidylinositol Biosynthesis Is a Homologue of Saccharomyces cerevisiae GPI2 Glycosylphosphatidylinositol (GPI) -anchors are used by many eukaryotic membrane proteins for ex’, 199, pp. 193–199.

Izquierdo, L., Acosta-Serrano, A., Mehlert, A. and Ferguson, M.A. (2015)

‘Identification of a glycosylphosphatidylinositol anchor-modifying β1-3 galactosyltransferase in Trypanosoma brucei’, *Glycobiology*, 25(4), pp. 438–447.

Izquierdo, L., Nakanishi, M., Mehlert, A., Machray, G., Barton, G.J. and Ferguson, M.A.J. (2009) ‘Identification of a glycosylphosphatidylinositol anchor-modifying β1-3 N-acetylglucosaminyl transferase in Trypanosoma brucei’, *Molecular Microbiology*, 71(2), pp. 478–491.

Jaquenoud, M., Pagac, M., Signorell, A., Benghezal, M., Jelk, J., Bütikofer, P. and Conzelmann, A. (2008) ‘The Gup1 homologue of Trypanosoma brucei is a GPI glycosylphosphatidylinositol remodelase’, *Molecular Microbiology*, 67(1), pp. 202–212.

Jenni, A., Knüsel, S., Nagar, R., Benninger, M., Häner, R., Ferguson, M.A.J., Roditi, I., Menon, A.K. and Bütikofer, P. (2021) ‘Elimination of GPI2 suppresses glycosylphosphatidylinositol GlcNAc transferase activity and alters GPI glycan modification in Trypanosoma brucei’, *Journal of Biological Chemistry*, 297(2), p. 100977.

Ji, Z., Tinti, M. and Ferguson, M.A.J. (2021) ‘Proteomic identification of the UDP-GlcNAc: PI α1–6 GlcNAc-transferase subunits of the glycosylphosphatidylinositol biosynthetic pathway of Trypanosoma brucei’, *PLOS ONE*. Edited by Z. Li, 16(3), p. e0244699.

Jones, N.G., Nietlispach, D., Sharma, R., Burke, D.F., Eyres, I., Mues, M., Mott, H.R. and Carrington, M. (2008) ‘Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein’, *Journal of Biological Chemistry*, 283(6), pp. 3584–3593.

Kabiri, M. and Steverding, D. (2021) ‘Molecular & Biochemical Parasitology Trypanosoma brucei transferrin receptor : Functional replacement of the GPI anchor with a transmembrane domain’, *Molecular & Biochemical Parasitology*, 242(November 2020), p. 111361.

Kajiwara, K., Watanabe, R., Pichler, H., Ihara, K., Murakami, S., Riezman, H. and Funato, K. (2008) ‘Yeast ARV1 Is Required for Efficient Delivery of an Early GPI Intermediate to the First Mannosyltransferase during GPI Assembly and Controls Lipid Flow from the Endoplasmic Reticulum’, *Molecular Biology of the Cell*. Edited by S. Munro, 19(5), pp. 2069–2082.

Käll, L., Krogh, A. and Sonnhammer, E.L.L. (2004) ‘A combined transmembrane topology and signal peptide prediction method’, *Journal of Molecular Biology*, 338(5), pp. 1027–1036.

Käll, L., Krogh, A. and Sonnhammer, E.L.L. (2007) ‘Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server’, *Nucleic Acids Research*, 35(SUPPL.2), pp. 429–432.

Kennedy, P.G.E. (2006) ‘Diagnostic and neuropathogenesis issues in human African trypanosomiasis’, *International Journal for Parasitology*, 36(5), pp. 505–512.

Kennedy, P.G.E. (2013) ‘Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness)’, *The Lancet Neurology*, 12(2), pp. 186–194.

Kennedy, P.G.E. and Rodgers, J. (2019) ‘Clinical and neuropathogenetic aspects of human African trypanosomiasis’, *Frontiers in Immunology*, 10(JAN), pp. 1–11.

Kingwell, K. (2021) ‘Pan-serotype antiviral for dengue infection’, *Nature Reviews Drug Discovery*, 20(4), pp. 263–263.

Kinoshita, T. (2014) ‘Biosynthesis and deficiencies of glycosylphosphatidylinositol’, *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 90(4), pp. 130–143.

Kinoshita, T. (2020) ‘Biosynthesis and biology of mammalian GPI-anchored proteins’, *Open biology*, 10(3), p. 190290.

Kinoshita, T. and Fujita, M. (2016) ‘Biosynthesis of GPI-anchored proteins: Special emphasis on GPI lipid remodeling’, *Journal of Lipid Research*, 57(1), pp. 6–24.

Kinoshita, T., Maeda, Y. and Fujita, M. (2013) ‘Transport of glycosylphosphatidylinositol-anchored proteins from the endoplasmic reticulum ☆’, *Biochim Biophys Acta .*, 1833(11), pp. 2473–2478.

Kruzel, E.K., Zimmett, G.P. and Bangs, J.D. (2017) ‘Life Stage-Specific Cargo Receptors Facilitate Glycosylphosphatidylinositol-Anchored Surface Coat Protein Transport in Trypanosoma brucei’, *mSphere*. Edited by A.P. Mitchell, 2(4), pp. 1–18.

Lake, J.A., De la Cruz, V.F., Ferreira, P.C.G., Morel, C. and Simpson, L. (1988) ‘Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences’, *Proceedings of the National Academy of Sciences of the United States of America*, 85(13), pp. 4779–4783.

Lane-Serff, H., MacGregor, P., Lowe, E.D., Carrington, M. and Higgins, M.K. (2014) ‘Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor’, *eLife*, 3, p. e05553.

Leidich, S.D., Kostova, Z., Latek, R.R., Costello, L.C., Drapp, D.A., Gray, W., Fassler, J.S. and Orlean, P. (1995) ‘Temperature-sensitive Yeast GPI Anchoring Mutants gpi2 and gpi3 Are Defective in the Synthesis of N-Acetylglucosaminyl Phosphatidylinositol.’, *Journal of Biological Chemistry*, 270(22), pp. 13029–13035.

Lillico, S., Field, M.C., Blundell, P., Coombs, G.H. and Mottram, J.C. (2003) 'Essential Roles for GPI-anchored Proteins in African Trypanosomes Revealed Using Mutants Deficient in GPI8', *Molecular Biology of the Cell*. Edited by R. Gilmore, 14(3), pp. 1182–1194.

Lindner, A.K., Lejon, V., Chappuis, F., Seixas, J., Kazumba, L., Barrett, M.P., Mwamba, E., Erphas, O., Akl, E.A., Villanueva, G., Bergman, H., Simarro, P., Kadima Ebeja, A., Priotto, G. and Franco, J.R. (2020) 'New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice', *The Lancet Infectious Diseases*, 20(2), pp. e38–e46.

Lukeš, J., Jirků, M., Doležel, D., Kral'ová, I., Hollar, L. and Maslov, D.A. (1997) 'Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic', *Journal of Molecular Evolution*, 44(5), pp. 521–527.

Macleod, O.J.S., Bart, J.M., MacGregor, P., Peacock, L., Savill, N.J., Hester, S., Ravel, S., Sunter, J.D., Trevor, C., Rust, S., Vaughan, T.J., Minter, R., Mohammed, S., Gibson, W., Taylor, M.C., Higgins, M.K. and Carrington, M. (2020) 'A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies', *Nature Communications*, 11(1), pp. 1–12.

Maeda, Y., Watanabe, R., Harris, C.L., Hong, Y., Ohishi, K., Kinoshita, K. and Kinoshita, T. (2001) 'PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER', *EMBO Journal*, 20(1–2), pp. 250–261.

Masterson, W.J., Doering, T.L., Hart, G.W. and Englund, P.T. (1989) 'A novel pathway for glycan assembly: Biosynthesis of the glycosyl-phosphatidylinositol anchor of the trypanosome variant surface glycoprotein', *Cell*, 56(5), pp. 793–800.

Masterson, W.J., Raper, J., Doering, T.L., Hart, G.W. and Englund, P.T. (1990) 'Fatty acid remodeling: A novel reaction sequence in the biosynthesis of trypanosome glycosyl phosphatidylinositol membrane anchors', *Cell*, 62(1), pp. 73–80.

Matthews, K.R. and Gull, K. (1994) 'Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes', *Journal of Cell Biology*, 125(5), pp. 1147–1156.

Mazhari-Tabrizi, R., Eckert, V., Blank, M., Müller, R., Mumberg, D., Funk, M. and Schwarz, R.T. (1996) 'Cloning and functional expression of glycosyltransferases from parasitic protozoans by heterologous complementation in yeast: The dolichol phosphate mannose synthase from Trypanosoma brucei brucei', *Biochemical Journal*, 316(3), pp. 853–858.

McDonough, V., Germann, M., Liu, Y., Sturley, S.L. and Nickels, J.T. (2002) 'Yeast

cells lacking the ARV1 gene harbor defects in sphingolipid metabolism: Complementation by human ARV1', *Journal of Biological Chemistry*, 277(39), pp. 36152–36160.

McKean, P.G. (2003) 'Coordination of cell cycle and cytokinesis in Trypanosoma brucei', *Current Opinion in Microbiology*, 6(6), pp. 600–607.

Medo, M., Aebersold, D.M. and Medová, M. (2019) 'ProtRank: Bypassing the imputation of missing values in differential expression analysis of proteomic data', *BMC Bioinformatics*, 20(1), pp. 1–12.

Mehlert, A. and Ferguson, M.A.J. (2007) 'Structure of the glycosylphosphatidylinositol anchor of the Trypanosoma brucei transferrin receptor', *Molecular and Biochemical Parasitology*, 151(2), pp. 220–223.

Mehlert, A., Richardson, J.M. and Ferguson, M.A.J. (1998) 'Structure of the Glycosylphosphatidylinositol Membrane Anchor Glycan of a Class-2 Variant Surface Glycoprotein from Trypanosoma brucei', 3.

Mehlert, A., Sullivan, L. and Ferguson, M.A.J. (2010) 'Glycotyping of Trypanosoma brucei variant surface glycoprotein MITat1.8', *Molecular and Biochemical Parasitology*, 174(1), pp. 74–77.

Mehlert, A., Wormald, M.R. and Ferguson, M.A.J. (2012) 'Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of trypanosoma brucei', *PLoS Pathogens*, 8(4), pp. 1–11.

Menon, A.K., Schwarz, R.T., Mayor, S. and Cross, G.A.M.M. (1990) 'Cell-free synthesis of glycosyl-phosphatidylinositol precursors for the glycolipid membrane anchor of Trypanosoma brucei variant surface glycoproteins. Structural characterization of putative biosynthetic intermediates', *Journal of Biological Chemistry*, 265(16), pp. 9033–9042.

Metcalf, P., Blum, M., Freymann, D., Turner, M. and Wiley, D.C. (1987) 'Two variant surface glycoproteins of Trypanosoma Brucei of different sequence classes have similar 6 Å resolution X-ray structures', *Nature*, 325(6099), pp. 84–86.

MILNE, K.G., FERGUSON, M.A.J. and MASTERSON, W.J. (1992) 'Inhibition of the GlcNAc transferase of the glycosylphosphatidylinositol anchor biosynthesis in African trypanosomes', *European Journal of Biochemistry*, 208(2), pp. 309–314.

Miyata, T., Takeda, J., Iida, Y., Yamada, N., Inoue, N., Takahashi, M., Maeda, K., Kitani, T. and Kinoshita, T. (1993) 'The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis', *Science*, 259(5099), pp. 1318–1320.

Montagna, G., Cremona, M.L., Paris, G., Amaya, M.F., Buschiazzo, A., Alzari, P.M.

and Frasch, A.C.C. (2002) ‘The trans-sialidase from the African trypanosome Trypanosoma brucei’, *European Journal of Biochemistry*, 269(12), pp. 2941–2950.
Montagna, G.N., Donelson, J.E. and Frasch, A.C.C. (2006) ‘Procyclic Trypanosoma brucei expresses separate sialidase and trans-sialidase enzymes on its surface membrane’, *Journal of Biological Chemistry*, 281(45), pp. 33949–33958.
Müller, L.S.M., Cosentino, R.O., Förstner, K.U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C.J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T.D., Saliba, A.-E., Sebra, R.P. and Siegel, T.N. (2018) ‘Genome organization and DNA accessibility control antigenic variation in trypanosomes’, *Nature*, 563(7729), pp. 121–125.
Murakami, Y., Siripanyaphinyo, U., Hong, Y., Tashima, Y., Maeda, Y. and Kinoshita, T. (2005) ‘The Initial Enzyme for Glycosylphosphatidylinositol Biosynthesis Requires PIG-Y, a Seventh Component’, *Molecular Biology of the Cell*, 16(11), pp. 5236–5246.
Nagamune, K., Nozaki, T., Maeda, Y., Ohishi, K., Fukuma, T., Hara, T., Schwarz, R.T., Sütterlin, C., Brun, R., Riezman, H. and Kinoshita, T. (2000) ‘Critical roles of glycosylphosphatidylinositol for Trypanosoma brucei’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(19), pp. 10336–10341.
Nagamune, K., Ohishi, K., Ashida, H., Hong, Y., Hino, J., Kangawa, K., Inoue, N., Maeda, Y. and Kinoshita, T. (2003) ‘GPI transamidase of Trypanosoma brucei has two previously uncharacterized (trypanosomatid transamidase 1 and 2) and three common subunits’, *Proceedings of the National Academy of Sciences of the United States of America*, 100(19), pp. 10682–10687.
Nett, I.R.E., Martin, D.M.A., Miranda-Saavedra, D., Lamont, D., Barber, J.D., Mehlert, A. and Ferguson, M.A.J. (2009) ‘The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness’, *Molecular and Cellular Proteomics*, 8(7), pp. 1527–1538.
Oberholzer, M., Morand, S., Kunz, S. and Seebeck, T. (2006) ‘A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes’, *Molecular and Biochemical Parasitology*, 145(1), pp. 117–120.
Okai, H., Ikema, R., Nakamura, H., Kato, M., Araki, M., Mizuno, A., Ikeda, A., Renbaum, P., Segel, R. and Funato, K. (2020) ‘Cold-sensitive phenotypes of a yeast null mutant of ARV1 support its role as a GPI flippase’, *FEBS Letters*, 594(15), pp. 2431–2439.
Orlean, P. and Menon, A.K. (2007) ‘Thematic review series : Lipid Posttranslational Modifications GPI anchoring of protein in yeast and mammalian cells , or : how we learned to stop worrying and love glycophospholipids’, *Journal Lipid Research*, 48(5),

pp. 993–1011.

Padilla-Mejia, N.E., Makarov, A.A., Barlow, L.D., Butterfield, E.R. and Field, M.C. (2021) 'Evolution and diversification of the nuclear envelope', *Nucleus*, 12(1), pp. 21–41.

Peacock, L., Bailey, M., Carrington, M. and Gibson, W. (2014) 'Meiosis and Haploid Gametes in the Pathogen Trypanosoma brucei', *Current Biology*, 24(2), pp. 181–186.

Pinger, J., Nešić, D., Ali, L., Aresta-Branco, F., Lilic, M., Chowdhury, S., Kim, H.S., Verdi, J., Raper, J., Ferguson, M.A.J.J., Papavasiliou, F.N. and Stebbins, C.E. (2018) 'African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat', *Nature Microbiology*, 3(8), pp. 932–938.

Pittet, M. and Conzelmann, A. (2007) 'Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1771(3), pp. 405–420.

Pontes De Carvalho, L.C., Tomlinson, S., Vandekerckhove, F., Jay Bienen, E., Clarkson, A.B., Jiang, M.S., Hart, G.W. and Nussenzweig, V. (1993) 'Characterization of a novel trans-sialidase of Trypanosoma brucei procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor', *Journal of Experimental Medicine*, 177(2), pp. 465–474.

Redmond, S., Vadivelu, J. and Field, M.C. (2003) 'RNAit: An automated web-based tool for the selection of RNAi targets in Trypanosoma brucei', *Molecular and Biochemical Parasitology*, 128(1), pp. 115–118.

Rico, E., Jeacock, L., Kovářová, J. and Horn, D. (2018) 'Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes', *Scientific Reports*, 8(1), pp. 1–10.

Rotureau, B. and Van Den Abbeele, J. (2013) 'Through the dark continent: African trypanosome development in the tsetse fly', *Frontiers in Cellular and Infection Microbiology*, 3(SEP), pp. 1–7.

Schägger, H. and von Jagow, G. (1991) 'Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form', *Analytical Biochemistry*, 199(2), pp. 223–231.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682.

Schwede, A., Macleod, O.J.S., MacGregor, P. and Carrington, M. (2015) 'How Does

the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins?', *PLoS Pathogens*, 11(12), pp. 1–18.

Segel, R., Aran, A., Gulsuner, S., Nakamura, H., Rosen, T., Walsh, T., Denda, H., Zeligson, S., Eto, K., Beeri, R., Okai, H., King, M.C., Levy-Lahad, E., Funato, K. and Renbaum, P. (2020) 'A defect in GPI synthesis as a suggested mechanism for the role of ARV1 in intellectual disability and seizures', *Neurogenetics*, 21(4), pp. 259–267.

Sharma, D.K., Smith, T.K., Crossman, A., Brimacombe, J.S. and Ferguson, M.A.J. (1997) 'acetylase of glycosylphosphatidylinositol membrane anchor biosynthesis in', 177, pp. 171–177.

Shimogawa, M.M., Saada, E.A., Vashisht, A.A., Barshop, W.D., Wohlschlegel, J.A. and Hill, K.L. (2015) 'Cell surface proteomics provides insight into stage-specific remodeling of the host-parasite interface in trypanosoma brucei', *Molecular and Cellular Proteomics*, 14(7), pp. 1977–1988.

Shishioh, N., Hong, Y., Ohishi, K., Ashida, H., Maeda, Y. and Kinoshita, T. (2005) 'GPI7 is the second partner of PIG-F and involved in modification of glycosylphosphatidylinositol', *Journal of Biological Chemistry*, 280(10), pp. 9728–9734.

Smith, T.K., Cottaz, S., Brimacombe, J.S. and Ferguson, M.A.J.J. (1996) 'Substrate specificity of the dolichol phosphate mannose: Glucosaminyl phosphatidylinositol $α$1-4-mannosyltransferase of the glycosylphosphatidylinositol biosynthetic pathway of African trypanosomes', *Journal of Biological Chemistry*, 271(11), pp. 6476–6482.

Sobering, A.K., Watanabe, R., Romeo, M.J., Yan, B.C., Specht, C.A., Orlean, P., Riezman, H. and Levin, D.E. (2004) 'Yeast Ras regulates the complex that catalyzes the first step in GPI-anchor biosynthesis at the ER', *Cell*, 117(5), pp. 637–648.

Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) 'Evolutionary diversity of eukaryotic small-subunit rRNA genes.', *Proceedings of the National Academy of Sciences*, 83(5), pp. 1383–1387.

Sternberg, J.M., Gierliński, M., Biéler, S., Ferguson, M.A.J. and Ndung'u, J.M. (2014) 'Evaluation of the Diagnostic Accuracy of Prototype Rapid Tests for Human African Trypanosomiasis', *PLoS Neglected Tropical Diseases*. Edited by J. Raper, 8(12), p. e3373.

Stevens, J.R. and Gibson, W. (1999) 'The molecular evolution of trypanosomes', *Parasitology Today*, 15(11), pp. 432–437.

Stokes, M.J., Murakami, Y., Maeda, Y., Kinoshita, T. and Morita, Y.S. (2014) 'New insights into the functions of PIGF, a protein involved in the ethanolamine phosphate

transfer steps of glycosylphosphatidylinositol biosynthesis', *Biochemical Journal*, 463(2), pp. 249–256.

Striegl, H., Andrade-Navarro, M.A. and Heinemann, U. (2010) 'Armadillo motifs involved in vesicular transport', *PLoS ONE*, 5(2).

Sullivan, L., Fleming, J., Sastry, L., Mehlert, A., Wall, S.J. and Ferguson, M.A.J. (2014) 'Identification of sVSG117 as an Immunodiagnostic Antigen and Evaluation of a Dual-Antigen Lateral Flow Test for the Diagnosis of Human African Trypanosomiasis', *PLoS Neglected Tropical Diseases*. Edited by J. Raper, 8(7), p. e2976.

Tanaka, S., Maeda, Y., Tashima, Y. and Kinoshita, T. (2004) 'Inositol Deacylation of Glycosylphosphatidylinositol-anchored Proteins Is Mediated by Mammalian PGAP1 and Yeast Bst1p', *Journal of Biological Chemistry*, 279(14), pp. 14256–14263.

Tinti, M., Güther, Maria Lucia S, Crozier, T.W.M., Lamond, A.I. and Ferguson, M.A.J. (2019) 'Proteome turnover in the bloodstream and procyclic forms of trypanosoma brucei measured by quantitative proteomics [version 1; peer review: 3 approved]', *Wellcome Open Research*, 4, pp. 1–26.

Tinti, M., Güther, Maria Lucia S., Crozier, T.W.M.M., Lamond, A.I. and Ferguson, M.A.J.J. (2019) 'Proteome turnover in the bloodstream and procyclic forms of trypanosoma brucei measured by quantitative proteomics [version 1; peer review: 3 approved]', *Wellcome Open Research*, 4, pp. 1–26.

Treumann, A., Güther, M.L.S., Schneider, P. and Ferguson, M.A.J. (2010) 'Analysis of the Carbohydrate and Lipid Components of Glycosylphosphatidylinositol Structures', in *Glycoanalysis Protocols*. New Jersey: Humana Press, pp. 213–236.

Treumann, A., Zitzmann, N., Hülsmeier, A., Prescott, A.R., Almond, A., Sheehan, J. and Ferguson, M.A.J. (1997) 'Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of Trypanosoma brucei', *Journal of Molecular Biology*, 269(4), pp. 529–547.

Turnock, D.C., Izquierdo, L. and Ferguson, M.A.J. (2007) 'The de novo synthesis of GDP-fucose is essential for flagellar adhesion and cell growth in Trypanosoma brucei', *Journal of Biological Chemistry*, 282(39), pp. 28853–28863.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', *Nature Protocols*, 11(12), pp. 2301–2319.

Urbaniak, M.D., Guther, M.L.S. and Ferguson, M.A.J. (2012) 'Comparative SILAC proteomic analysis of trypanosoma brucei bloodstream and procyclic lifecycle stages', *PLoS ONE*, 7(5).

Urwyler, S., Studer, E., Renggli, C.K. and Roditi, I. (2007) 'A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei', *Molecular Microbiology*, 63(1), pp. 218–228.

Vanhollebeke, B., De Muylder, G., Nielsen, M.J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S.K. and Pays, E. (2008) 'A Haptoglobin-Hemoglobin Receptor Conveys Innate Immunity to Trypanosoma brucei in Humans', *Science*, 320(5876), pp. 677–681.

Vidugiriene, J. and Menon, A.K. (1994) 'The GPI anchor of cell-surface proteins is synthesized on the cytoplasmic face of the endoplasmic reticulum', *Journal of Cell Biology*, 127(2), pp. 333–341.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O. and Vázquez-Baeza, Y. (2020) 'SciPy 1.0: fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272.

Vol, N., Training, M.S., Investigator, V.A.M. and Dnase, M. (1982) '- .. • - - .- .- 2.1', 299(September), pp. 451–453.

Wall, R.J., Rico, E., Lukac, I., Zuccotto, F., Elg, S., Gilbert, I.H., Freund, Y., Alley, M.R.K., Field, M.C., Wyllie, S. and Horn, D. (2018) 'Clinical and veterinary trypanocidal benzoxaboroles target CPSF3', *Proceedings of the National Academy of Sciences of the United States of America*, 115(38), pp. 9616–9621.

Wang, H., Lockwood, S.K., Hoeltzel, M.F. and Schiefelbein, J.W. (1997) 'The ROOT HAIR DEFECTIVE3 gene encodes an evolutionarily conserved protein with GTP-binding motifs and is required for regulated cell enlargement in Arabidopsis', *Genes and Development*, 11(6), pp. 799–811.

Wang, Y., Maeda, Y., Liu, Y.S., Takada, Y., Ninomiya, A., Hirata, T., Fujita, M., Murakami, Y. and Kinoshita, T. (2020) 'Cross-talks of glycosylphosphatidylinositol biosynthesis with glycosphingolipid biosynthesis and ER-associated degradation', *Nature Communications*, 11(1), pp. 1–18.

Watanabe, R., Inoue, N., Westfall, B., Taron, C.H., Orlean, P., Takeda, J. and Kinoshita, T. (1998) 'The first step of glycosylphosphatidylinositol biosynthesis is mediated by a complex of PIG-A, PIG-H, PIG-C and GPI1', *EMBO Journal*, 17(4), pp. 877–885.

Watanabe, R., Kinoshita, T., Masaki, R., Yamamoto, A., Takedat, J. and Inoue, N. (1996) 'PIG-A and PIG-H, which participate in glycosylphosphatidylinositol anchor biosynthesis, form a protein complex in the endoplasmic reticulum', *Journal of Biological Chemistry*, 271(43), pp. 26868–26875.

WHO (2021) *Ending-the-neglect-to-attain-the-SDGs--NTD-Roadmap*.

Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2007) 'The SUPERFAMILY database in 2007: Families and functions', *Nucleic Acids Research*, 35(SUPPL. 1), pp. 308–313.

Wirtz, E., Leal, S., Ochatt, C. and Cross, G.A.M.M. (1999) 'A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei', *Molecular and Biochemical Parasitology*, 99(1), pp. 89–101.

Ziegelbauer, K. and Overath, P. (1992) 'Identification of invariant surface glycoproteins in the bloodstream stage of Trypanosoma brucei', *Journal of Biological Chemistry*, 267(15), pp. 10791–10796.

Zoltner, M., Leung, K.F., Alsford, S., Horn, D. and Field, M.C. (2015) 'Modulation of the Surface Proteome through Multiple Ubiquitylation Pathways in African Trypanosomes', *PLoS Pathogens*, 11(10), pp. 1–26.

Supplementary Data

Appendix A

Preliminary TbGPI12-Myc IP experiment using PCF cells.

Samples of 6×10^8 PCF WT and TbGPI12 conditional null mutant Tet induced cells were lysed with either 0.5% digitonin (Fig S1A) or 1% TX-100 (Fig S1B) lysis buffer and IP was performed to the cell lysates with Myc-Trap™ beads as described in section 4.1.3 and proteins were eluted with sodium citrate (pH 2.8), separated on SDS-PAGE and silver stained with Invitrogen SilverQuest™. Although TbGPI12-Myc protein showed a strong signal when detected with rat anti-Myc antibody (Fig 4.3), there was no obvious corresponding silver stained band for TbGPI12-Myc at 28.6 kDa in the IPs (Fig S1A and B, lane 2), nor for its putative partner protein TbMTI. Nevertheless, the regions (red rectangles) that were expected to contain TbMTI (upper band) and TbGPI12 (lower band) were cut and send for proteomics. TbGPI12 was identified in both the 0.5% digitonin and 1% TX-100 IP samples with protein coverage of 15% and 11%, respectively but there was no identification for TbMTI from the upper bands.

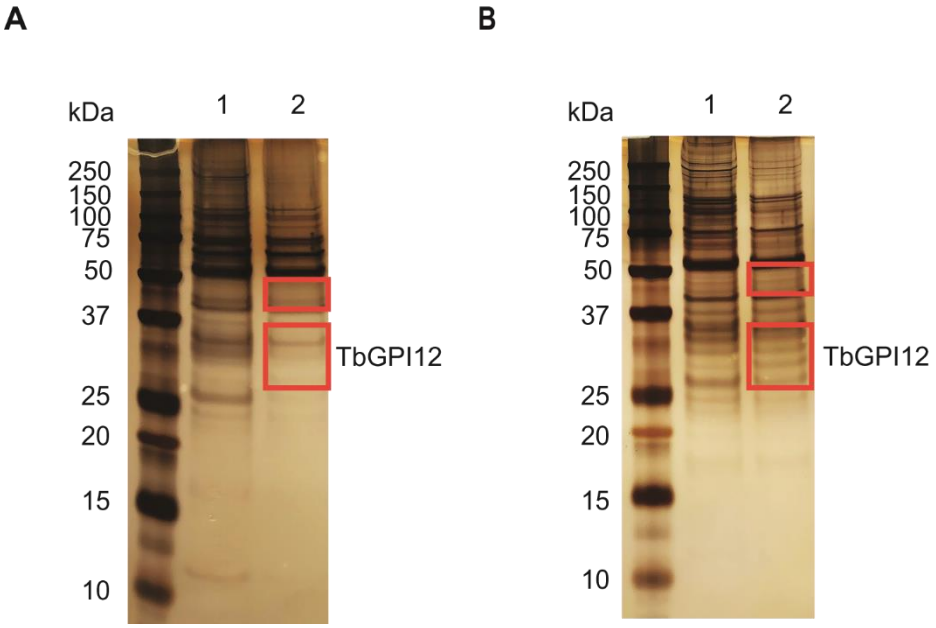

Figure S1 Silver staining of SDS-PAGE of TbGPI12-Myc IP products.

A: Cell lysates (0.5% digitonin) from the PCF WT (lane 1) or TbGPI12 conditional null mutant Tet induced (lanes 2) were subjected to IPs, and the proteins were eluted with sodium citrate (pH2.8). The eluted proteins were then resolved on SDS-PAGE, and silver staining was applied with Invitrogen set kit. The regions expected to contain TbMTI and TbGPI12 (labelled with red

rectangles) were subjected to proteomics identification. B: Duplicate experiments but with 1%TX-100 as cell lysis condition.

Appendix B

In situ tagging and detection of TbMTI

In this experiment, we used *in situ* C-terminal tagging of TbMTI with 3 × c-Myc using the pMOTag plasmid (Oberholzer *et al.*, 2006) described in section 5.1. Tagging in these clones was checked by immunoblotting; BSF TbMTI-3Myc expressing cell lines were lysed with 1%TX-100 followed by pull-down with MycTrap™ beads, proteins were eluted from the beads with SDS sample buffer and detected by anti-Myc Western blot (Fig S2). Even with 1×10^9 cells as input we could only detect a trace amount of TbMTI-3Myc. To date, the *in situ* tagging of TbMTI in PCF *T. brucei* hasn't been successful and we have paused studies using TbMTI as bait protein.

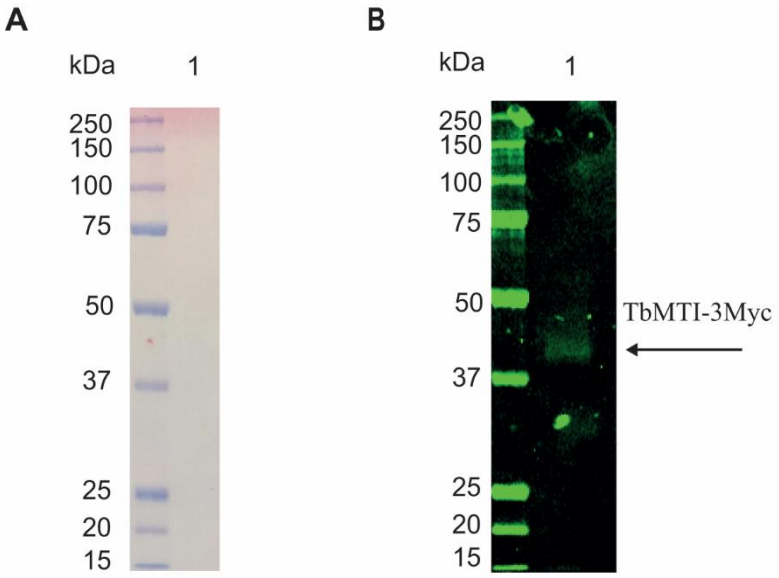

Figure S2 Immunoprecipitation and Western blot of TbMTI-3Myc.

A and B: Ponceau staining and anti-Myc Western blot of TbMTI-3Myc IP products. 1×10^9 TbMTI-3Myc expressing cells (lane 1) were lysed with 1%TX-100 and IP was performed on the supernatant with Myc-Trap™ beads. Proteins were eluted with SDS sample buffer and separated on SDS-PAGE. After transferring to a nitrocellulose membrane, the blot was developed with a rat anti-Myc antibody.